Validating Stereoscopic Volume Rendering

David Roberts

A Thesis presented for the degree of

Doctor of Philosophy

Innovative Computing Group
School of Engineering and Computing Sciences
Durham University
United Kingdom

2016

Dedicated to

My Mom and Dad, my sisters, Sarah and Alison, my niece Chloe, and of course Keeta.

Validating Stereoscopic Volume Rendering

David Roberts

Submitted for the degree of Doctor of Philosophy

2016

Abstract

The evaluation of stereoscopic displays for surface-based renderings is well established in terms of accurate depth perception and tasks that require an understanding of the spatial layout of the scene. In comparison direct volume rendering (DVR) that typically produces images with a high number of low opacity, overlapping features is only beginning to be critically studied on stereoscopic displays. The properties of the specific images and the choice of parameters for DVR algorithms make assessing the effectiveness of stereoscopic displays for DVR particularly challenging and as a result existing literature is sparse with inconclusive results.

In this thesis stereoscopic volume rendering is analysed for tasks that require depth perception including: stereo-acuity tasks, spatial search tasks and observer preference ratings. The evaluations focus on aspects of the DVR rendering pipeline and assess how the parameters of volume resolution, reconstruction filter and transfer function may alter task performance and the perceived quality of the produced images.

The results of the evaluations suggest that the transfer function and choice of reconstruction filter can have an effect on the performance on tasks with stereoscopic displays when all other parameters are kept consistent. Further, these were found to affect the sensitivity and bias response of the participants. The studies also show that properties of the reconstruction filters such as post-aliasing and smoothing do not correlate well with either task performance or quality ratings.

Included in the contributions are guidelines and recommendations on the choice of parameters for increased task performance and quality scores as well as image based methods of analysing stereoscopic DVR images.

Declaration

The work in this thesis is based on research carried out at the Innovative Computing Group, the School of Engineering and Computing Sciences, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

First, my deepest gratitude goes to my parents Pat and Anthony Roberts, my sisters Sarah and Alison and my niece Chloe. Quite simply without their support I would not have managed the last few years. Thanks for ensuring I'm always part of the family.

I would like to express thanks to my supervisor Dr Ioannis Ivrissimtzis for his advice and continual support throughout the PhD and for the many hours of his time. I would also like to thank Prof. Nick Holliman for supervising me during my initial year of studies. Funding for this PhD was provided by the EPSRC.

Publications

Research in the thesis has been document in part in the following publications:

"Investigating Depth Perception with Stereoscopic Volume Rendering"
D.A.T. Roberts. Proceedings of EG UK Computer Graphics and Visual Computing. September 2014.

"An Evaluation of Reconstruction Filters for a Path-Searching Task in 3D"
D.A.T. Roberts, I. Ivrissimtzis and N.S. Holliman. Sixth International Workshop on Quality of Multimedia Experience (QoMEX). September 2014.

"Quality Measures of Reconstruction Filters for Stereoscopic Volume Rendering"
D.A.T. Roberts, I. Ivrissimtzis. Computational Visual Media 2(1). April 2016.

"Reevaluating Reconstruction Filters for Path-Searching Tasks in 3D"
D.A.T. Roberts, I. Ivrissimtzis. Computer Graphics Forum (Early Access). July 2016.

Further research conducted in part by the author has been documented in:

"3D Sound and Image Interactions: A Review of Audio-Visual Depth Perception"
J.S. Berry, D.A.T. Roberts and N.S. Holliman. Human Vision and Electronic Imaging XIX. SPIEE The International Society for Optical Engineering 9014. February 2014.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis the emerging field of stereoscopic volume rendering is explored. Aspects of the volume rendering pipeline are analysed by the use of spatial awareness tasks, quality measures and image processing techniques. The results found in the thesis are applicable to an array of fields that use volume rendering that includes the medical domain, computational fluid dynamics and flight simulations.

The research is based on four experiments. First, an experiment is performed to determine how different volume resolutions can affect fine depth perception of participants using a stereo-acuity test. Following this a second experiment is performed to assess what impact the choice of reconstruction filter used in the Direct Volume Rendering (DVR) pipeline has on the performance of a path-searching task. The scope of the evaluation is then extended to include a wider range of filters with different properties.

Further, we introduce novel analysis methods using techniques from the signal detection theory literature. Finally, we analyse the subjective ratings of reconstruction filters used to generate images on four different quality measures.

This chapter will provide the context for evaluating stereoscopic volume rendering as well as defining the problem the thesis seeks to address. A brief background of direct volume rendering and stereoscopy will be given followed by the aims and objectives. The chapter concludes with an overview of the thesis and research contributions.

1.1 Background

1.1.1 Direct Volume Rendering

Direct volume rendering is a technique used to visualise volumetric data. Volumetric data are commonly represented as a lattice of scalar values acquired from scanning devices such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), computer simulations and by converting standard surface geometry. Typically the data does not contain any well-defined surfaces and so if standard polygon rendering methods are used large amounts of data in the volume may not be present in the final image. Direct volume rendering allows the entire volume to be visualised in a single image by giving the viewer control over the transparency and colour of features in the volume. Figure 1.1 shows three different volumes rendered using direct volume rendering.

The classical DVR algorithm casts a ray for each pixel in the image into the volumetric data. At sample points along each ray the opacity and colour for that point are computed. Due to the discrete nature of the volumetric data, interpolation is required when samples lie on off-lattice positions. A reconstruction filter is used for this purpose.

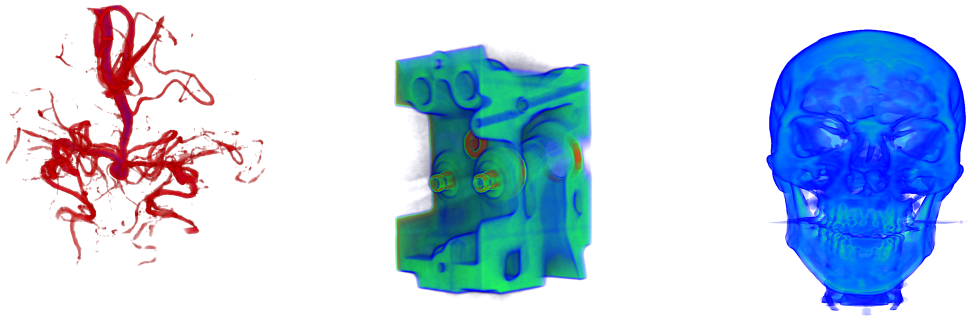
There are a variety of parameters in the algorithm that can affect what aspects of the volume are displayed as well as other parameters that can have an impact on the visual quality of the produced images. This includes the number of samples to take when sampling, the reconstruction filter used, as well as the transfer function to assign colours and opacities to the volume.

The range of parameters and the complex nature of the volumetric data itself can lead to visualisations that are visually difficult to comprehend. The mathematical background to DVR as well as details regarding reconstruction filters are elaborated in Chapter 2.

1.1.2 Stereoscopy

The human visual system (HVS) sees two overlapping views of the world, the horizontal disparity between matching points in the views creates a binocular visual field. This gives rise to the stereoscopic depth cue that allows an observer to perceive a single 3D view of a scene.

By showing two horizontally displaced images of the same scene, 3D images can be displayed using either stereoscopic displays or standard 2D displays using anaglyph images. Figure 1.2 shows an example of a stereoscopic image rendered using the anaglyph



(a) Angiography (b) Engine (c) Visible Human

Figure 1.1: Direct volume renderings of three different volume data sets. The angiography and engine datasets are publicly available at http://www.volvis.org. The Visible Human dataset is provided by the National Library of Medicine (Spitzer & Whitlock 1998)

technique.

Stereoscopic display of images can aid in depth judgements as well as assisting the understanding of how scenes are organised spatially. The accuracy of tasks such as path-searching, where paths must be located in an image, have also found to be improved when the scene is viewed stereoscopically. However, stereoscopic images are susceptible to 2D image artefacts that can affect the quality of a 3D image. Aliasing in 3D images can lead to inaccurate positions and sizes of objects in a scene. Further, when stereoscopy is combined with the perspective depth cue, aliasing can lead to depth positioning errors (Pfautz 2000). Other objects such as blurring can decrease the ability for observers to perceive fine differences in depth between two objects (Costa et al. 2010). Chapter 3 discusses in detail the perception of depth including the properties of different cues and aspects that can affect how depths are understood.

1.2 Context and Problem Definition

Volumetric data used in DVR can describe complex objects that do not have a simple layout. This can include overlapping spatial features as well as fine details. The choice of transfer function used in DVR algorithms can also lead to a large number of semi-transparent features in the final image. Further, parameters that must be decided before producing an image, including the sampling strategy and reconstruction filter, can produce various other visual artefacts.

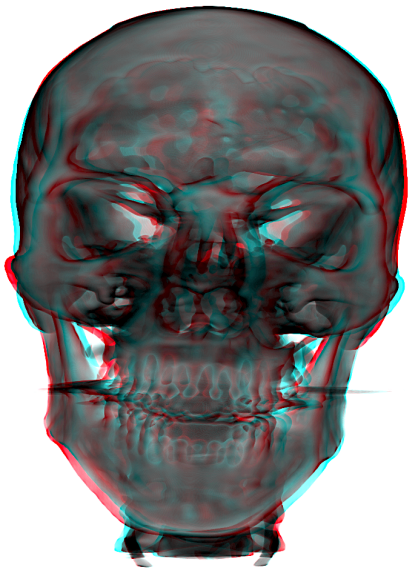


Figure 1.2: Monochrome anaglyph stereoscopic rendering of the Visible Human data set (Spitzer & Whitlock 1998). Requires red/cyan glasses for 3D viewing.

Several methods have been used to aid in the understanding of depth in DVR by using monocular depth cues. These include shading, shadowing, feature halos to surround regions of interest and 2D pictorial depth cues. A concern with these methods is that in order to increase the perception of depth the DVR image is altered in some way which can obscure details in the scene.

As an alternative, stereoscopic displays can be used to to introduce the binocular depth cue to DVR scenes without altering the colours or clarity of the DVR image. For surface-based renderings, stereoscopic displays have been found to aid in depth judgements, interpreting surface curvature and spatial localization (Holliman 2005). Regarding DVR images, stereoscopic rendering can improve the results of tasks, such as depth ordering in (Cho et al. 2014), however the results from the literature are neither unanimous or consistent. The main challenge is that a variety of parameters must be addressed when generating stereoscopic images that includes camera separation, viewing distance as well as properties unique to each stereoscopic display and this list of parameters is extended when generating DVR images.

This thesis will address the problems of whether stereoscopic displays can aid in tasks

requiring depth perception compared to standard 2D displays. Further, and critically, whether parameters of the DVR pipeline can alter the performance of tasks performed in 3D and whether the results relate to properties of the parameters. Following this the thesis will investigate how a single parameter, the choice of reconstruction filter, can affect the perceived quality of the stereoscopic DVR images. From this a compactness image based metric will be established and correlated with the perceived quality ratings of the DVR images.

1.3 Hypothesis

Different aspects of stereoscopic volume rendering are explored in this thesis, however there is a single overarching theme of analysing the effectiveness of DVR in task based and subjective quality settings. The overall hypothesis for this thesis is therefore that the parameters and settings of the DVR algorithm, including depth resolution, opacity and reconstruction filter, can affect the performance of spatial tasks when using DVR as well as the perceived quality of depth. Further, that the differences in results between these parameters are significant enough that producers of DVR images should be aware of the impact the parameters can have.

1.4 Aims and Objectives

This thesis aims to investigate aspects of stereoscopic volume rendering with a particular focus on stages of the volume rendering pipeline for tasks requiring depth perception. The objectives of the research are listed below.

1. Investigate existing direct volume rendering algorithms with a focus on artefacts and the effects of different parameter choices in the algorithms.

2. Develop and run an experiment to investigate the effect the resolution of the volume data has on fine depth perception. This is reflected in Chapter 4.

3. Assess whether the choice of reconstruction filter affects a task requiring depth perception. Evaluate a range of filters and assess whether measurable properties of the reconstruction filters and images relate to task performance. This is achieved in Chapter 5 and extended in Chapter 6.

4. Assess the impact of the type of transfer function on task performance. This is analysed in Chapter 6.

5. Assess how the reconstruction filter may affect the perceived quality of direct volume rendered images. This has been assessed in Chapter 7.

6. Provide guidelines on the choice of parameters to be used with stereoscopic DVR based on the results of the prior objectives.

1.5 Limitations

There are limitations to the research reported in this thesis, these are given below and detailed in each chapter.

1. The research focuses on stereoscopic depth perception and therefore depth cues such as motion parallax and the kinetic depth effect are not studied.

2. Only novice participants have been used for the experiments. The choice of domain or expert users is an open topic in the literature.

3. In order to control the size of the experiments a limited set of data sets have been used to produce the DVR images used in the experiments.

1.6 Thesis Overview and Research Contributions

The summaries and main contributions of each of the chapters contained within this thesis are given below. Due to the thesis being an intersection between DVR and stereoscopy there are two background chapters, Chapter 2 and Chapter 3 that provide the relevant background to each topic respectively. Chapters 4, 5, 6 and 7 are the content chapters reflecting the research conducted.

Chapter 2 - Volume Rendering

This chapter begins with an overview of volume rendering algorithms and common methods for acquiring volumetric data. The mathematical background of DVR is then detailed with a particular focus on the sampling and reconstruction stages of the DVR pipeline. Brief descriptions of practical DVR implementations are then given.

Chapter 3 - Human Vision and Depth Perception

In this chapter details of the human visual system are introduced with an emphasis on the perception of depth. This is followed by an overview of 3D displays including technical considerations. The chapter concludes with a discussion of factors that can affect depth perception including forms of aliasing and blurring.

Chapter 4 - Investigation of Depth Perception with Stereoscopic Volume Rendering

In this chapter the results of an experiment investigating the finest depth that can be perceived with stereoscopic DVR are reported. We test how different resolutions of volumetric data may affect the results of a modified stereo-acuity test using a staircase design to find the depth thresholds. The results show that the resolution of the volume does not have a significant effect on fine depth perception but high resolution volumes can increase response time. Work from this chapter has been published in Roberts, Investigating Depth Perception with Stereoscopic Volume Rendering (Roberts 2014).

Chapter 5 - An Evaluation of Reconstruction Filters for a Path-Searching Task in 3D

This chapter presents an initial experiment to examine how the choice of reconstruction filter used in DVR can affect the results of a path-searching task with 3D displays. Three reconstruction filters are assessed and a computer generated data set is used that has some properties of natural data sets. The results suggest that as the quality of reconstruction filter increases the accuracy of the task also increases. Work from this chapter has been published in Roberts, Ivrissimtzis and Holliman, An Evaluation of Reconstruction Filters for a Path-Searching Task in 3D (Roberts et al. 2014).

Chapter 6 - Reevaluating Reconstruction Filters for a Path-Searching Task in 3D

In this chapter we extend the analysis of reconstruction filters for a path-searching task to include five reconstruction filters each with two different transfer functions in monoscopic and stereoscopic display conditions. We seek to determine how the post-aliasing and smoothing properties of the filters relate to task performance. Further, techniques from signal detection theory are applied to analyse the results. The

results suggest that post-aliasing and smoothing have no simple correlation to task accuracy and that reconstruction filters have the most significant effect on the bias and sensitivity of participants towards the task. Work from this chapter has been submitted to Computer Graphics Forum, Roberts and Ivrissimtzis, Reevaluating Reconstruction Filters for a Path-Searching Task in 3D (Roberts & Ivrissimtzis 2016*b*).

Chapter 7 - Quality Measures of Reconstruction Filters for Stereoscopic Volume Rendering

In this chapter we analyse participants' ratings of stereoscopic volume rendered scenes on four subjective quality measures: depth quality, depth layout, minimal jaggyness and sharpness. The relationship between the properties of the reconstruction filters, post-aliasing and smoothing, and the quality measures are assessed and found to be statistically insignificant. However, we find that the image specific measure of shape compactness has a high correlation with each of the quality measures. The results suggest that the B-spline approximation filter should be preferred for higher perceived quality. Work from this chapter has been published in Roberts and Ivrissimtzis, Quality Measures of Reconstruction Filters for Stereoscopic Volume Rendering (Roberts & Ivrissimtzis 2016*a*).

Chapter 8 - Conclusions and Future Work

In Chapter 8 a review of the research presented is given as well as summaries of the results and novel contributions. The chapter concludes with possible directions for future work.

Chapter 2

Volume Rendering

2.1 Introduction

This chapter will provide background on volume rendering. The theory common to all methods will be explained and specific volume rendering algorithms will be detailed. Artefacts caused by choices in the algorithms will be investigated with particular interest in errors caused by sampling and non-ideal reconstruction filters. These sections are included due to their usage when analysing sampling strategies and reconstruction filters used in the experiments in the research chapters.

2.2 Indirect Volume Rendering versus Direct Volume Rendering

Volume rendering algorithms fall under two main categories, Indirect Volume Rendering (IVR) also known in the literature as Surface Fitting (SF) and Direct Volume Rendering (DVR) depending on the usage of proxy geometry.

2.2.1 Indirect Volume Rendering

The initial approach to rendering volumetric data, IVR, converted the data into a set of iso-surfaces, the classic approach being the Marching Cubes algorithm (Lorensen & Cline 1987). These surfaces were then rendered using standard polygon based graphics hardware. This approach partitions the data set into cells and determines the possible ways in which the surface may interact with each cell. From this triangles are generated that approximate the surface of the volume.

Although potentially faster than DVR algorithms due to the use of standard geometry that can be rendered natively on graphics hardware, a general issue with this category is that the data set must be converted to the surface form and so a binary decision must be made as to whether a sample point is or is not included in the surface. If the sample point is not included in a surface then the point will not be rendered. This can make it particularly difficult to render small features as well as phenomena that do not readily fit a surface, such as fog (Kaufman & Mueller 2005). The binary segmentation of the surfaces in the data can also lead to representations that do not match the actual dataset. To create more accurate representations larger number of geometric primitives are required that can overwhelm the capability of the graphics hardware. As only the surface of a volume is rendered a potentially large amount of useful information contained within the volume is discarded. A secondary problem related to the surface-based rendering is that the density of a volume is ignored. Objects that are the same but have a different thickness will be rendered identically.

2.2.2 Direct Volume Rendering

DVR does not attempt to fit the data to a set of surfaces, rather it renders the data directly to the image plane, an example is the raycasting algorithm (Levoy 1988). These methods do not limit the visualisation to just a set of surfaces and so amorphous data sets such as fog can be rendered. A secondary advantage of volumetric rendering is the ability to perform binary operations on the volume. This allows subsets and the removal of slices of the volume without the necessity for expensive recalculation that is required for surface based rendering.

The disadvantage of the DVR approach is a potential increase in processing time (Meißner et al. 2000, Elvins 1992) and a suggestion of ambiguity due to multiple sample points contributing to a single pixel (Boucheny et al. 2009). This thesis will furthermore focus on DVR.

2.3 Data

The volume data to be rendered takes the form of a grid lattice where each (x_i, y_i, z_i) holds an array of some scalar values $f_0(x_i)$ (Levoy 1988). In the case that there are multiple values per location, then this is known as a multivariate data set. The scalar

values in the data set represent a particular property of the original stimulus that can include density, heat or some other measurable property. There are two ways the scalar values can be treated on a grid lattice, the first is the value is treated as a voxel and is constant over the area containing the sample point. This is analogous to a pixel in a 2D image. The second method involves treating each value as a sample point lying on a gridpoint. In this thesis we will use the latter interpretation. These values themselves can be acquired from a variety of different sources, primarily from the medical domain and disparate scanning devices but also from simulations and algebraic geometry. Different methods of data acquisition are briefly described.

2.3.1 Medical Devices

Computed Tomography (CT) is a common scanning process that makes use of X-Rays that traverse through a patient and are recorded on the other side once they have passed through the body. The X-rays are attenuated differently depending on the materials within the body, the different attenuation rates can then be used to classify the different properties of the body. 2D cross-sections of the stimulus are generated by rotating the X-ray emitting device around the patient and moving laterally to produce a final 3D cross-section. This can then be processed into a volumetric data set. A particular problem with computed tomography is the use of ionising radiation to produce the X-rays. There is a potential for a slight increase in risk of cancer depending on dosage and number of scans (Brenner & Hall 2007). Due to health risks involved with high doses of radiation, typically the resolutions of the data sets acquired are lower than other modularities.

In terms of data quality a number of artefacts can occur in the data and much work has taken place to identify and rectify the errors (Boas & Fleischmann 2012). Noise is one of the problems that can occur in the process of scanning and is a result of statistical errors in low photon counts used in the X-rays that cause bright and dark thin streaks to appear in the data. Motion artefacts can also be caused by a patient moving whilst the scanning is taking place. This results in blurring and ghosting with streaks near edges of high contrast. This artefact can be lessened somewhat by reducing the time required for the scan thereby giving the patient less time to move.

Magnetic Resonance Imaging (MRI) uses the theory that nuclei in an atom emit radiation when placed within a magnetic filed of sufficient strength. Different particles in the source stimulus will emit unique forms of radiation allowing different materials to be

detected. A significant difference between MRI and CT scanning devices is that MRI does not produce ionising radiation and therefore can potentially create larger data resolutions. Similar to CT data, MRI can suffer from the movement of the patient as well as variations in the magnetic field that can cause the data to be mispositioned within the image (Pickens 2000).

Other devices capable of producing volumetric data include *ultrasound* imaging systems and *Positron Emission Tomography* (PET). The development of 2D ultrasound images dates back to 1952 with the work done by John J. Wild (1952). It is a non-destructive process that directs high frequency sound waves into the stimulus. The different tissues within the stimulus will return different echoes of the sound and by recording these and classifying them a 2D image can be produced. Volumetric data can be generated by combining multiple 2D ultrasound images. The scanning procedures described above assess the anatomical features of the stimulus in terms of size, density and shape. However, PET can assess the functional properties of the stimulus being scanned, for example it can be used to determine the inner behaviour of a tumor. The PET system detects positrons emitted by a tracer element introduced into the body. By tracking the movement of the tracer element volume data on metabolic processes of the body can be produced (Juweid & Cheson 2006, Ollinger & Fessler 1997).

2.3.2 Voxelization

Via a process known as *voxelization*, or 3D scan-conversion, volumetric data can be generated from other geometric representations of the object. This representation allows objects defined by standard polygon meshes to interact with and combine in a standard volume rendered scene. A set of voxels are generated with the highest quality algorithms producing the most accurate discrete approximation of the continuous geometric object. The accuracy of the discrete representation depends upon the algorithm and to some extent the resolution of the data set.

Initial algorithms such as the method presented by Kaufman (1987) used what is referred to as *binary* or point sampling voxelization. For each voxel in the desired resolution if it is contained within the geometry then it is assigned the value 1, else it is given the value 0. Whilst the models generated are consistent, such a process leads to what is referred to as object space aliasing. This is because each point is either inside or outside the volume, no intermediary values are allowed and therefore the quality of the final rendered image

depends entirely upon the resolution of the voxelization. Low resolution data sets that exhibit object space aliasing will lead to jagged edges and discontinuities in the final image.

Although higher resolutions will reduce the overall impact of object space aliasing, this will increase the memory required to store the volume. These issues can be alleviated by performing a low-pass filter that removes high frequencies that represent sharp edges on the geometric object and then sampling the scene. The filtering acts as an averaging process with the precise value being determined by a weighted spherical filter (Wang & Kaufman 1993, 1994). The intermediary values produce a volume that does not have well-defined boundaries between objects eliminating the object space aliasing artefacts of the prior algorithms. A secondary benefit of this form of voxelization is that when a Gaussian filter is used to produce the values the resulting data set matches closely to data sets that can be produced from real scanning devices. This is because real scanning devices have a point spread function that is not a dimensionless sample point and the value is taken as an average over some finite area (Sramek & Kaufman 1998).

Distance field techniques introduced to volume rendering in Breen et al. (1998) and extensively reviewed in Jones et al. (2006) are an alternative to the filtering techniques of voxelization. A distance volume is generated where each voxel is assigned a value depending on the shortest distance to a surface point of an object in the scene. Points on the surface of an object are assigned the value 0, to distinguish points that lie inside or outside of an object values inside an object are given a negative value and points outside an object are given a positive value.

The distance field approach has two advantages over filtering for voxelisation. Firstly the filtering method must incorporate the use of distance fields. Therefore by sampling the distance fields directly a layer of complexity can be removed. Secondly distance fields ensure an accurate reconstruction of iso-surfaces in the volume.

Although capable of higher quality voxelisation than filtering methods, there are two conditions that must be guaranteed when using distance fields. This is primarily because of the unavoidable limits imposed by the sampling theorem. The first condition is that the curvature of the objects to be voxelised should be low in relation to the resolution of the volume that will be created thereby reducing reconstruction errors. Secondly the reconstruction and gradient filters should not use samples that are distributed on both sides of a medial surface of the object or on the complement of the object. A medial surface in this respect is a set of points that are equidistant to a minimum of two points

on the boundary surface of the object. This condition is required as the gradient at these points will be undefined. Therefore these points should not be used by the gradient filter if it is to be used for shading when producing the final rendering.

Distance fields can be computationally expensive to compute, and therefore an alternative to using direct distance fields is by using vector methods. A vector to the nearest surface is stored in each voxel, these are then propagated to the neighbouring voxels over a number of passes before a final distance value is calculated. Whilst this method is slightly less accurate than the true distance field calculation it is significantly faster to process (Jones & Satherley 2000).

The voxelisation method proposed by Novotny et al. (2010) uses the distance fields approach by using a two-pass method. The first pass identifies areas that contain sharp details that will need modification later. The second pass modifies the previously identified areas so as to round the sharp details and determine an appropriate curvature for the given volume resolution. Truncated distance fields are then used to decrease the memory required when higher volume resolutions are used. Truncated distance fields only store values in the vicinity of the surfaces with the inside and outside values being stored separately.

Recent advances in the field of programmable graphics processors have allowed voxelization to be performed interactively and so dynamic scenes can be generated, voxelized and rendered in real-time (Schwarz & Seidel 2010, Pantaleoni 2011)

2.3.3 Sampling Lattices

There are several possibilities for how data samples may be arranged in the 3D volume which generally fall under two categories *structured* and *unstructured*. Structured layouts have an underlying strategy to the data set and within this category there are three main lattice types, Cartesian Cubic (CC), Body-Centred Cubic (BCC) and Face-Centred Cubic (FCC). CC lattices are the most common due to ease of use and that most software assumes this format. In this case the grid lattice is rectilinear with distances d_x, d_y, d_z between each data sample per axis. Regular rectilinear, or *isotropic*, data sets are a special case of rectilinear CC grids where the distance between each sample are equal in each case so $d_x = d_y = d_z$. Due to the Cartesian lattice, reconstruction filters that can be applied to 1D data sets are straightforward to generalise to 3D which makes this a particularly convenient format. The CC lattice, although simple to use and understand, is not the

most efficient for storing band-limited data sets. Efficiency in this case is measured by how few sample points are required to be stored without suffering artefacts caused by under sampling the data set.

The BCC lattice can be thought of as a CC lattice with an extra sample point in the centre of each cell. When a signal is band-limited and sampled at the Nyquist rate the BCC layout produces an optimal sampling strategy requiring only 71% of the number of sample points of the CC lattice for the same accuracy.

As an alternative to the CC and BCC lattices, the FCC lattice can be visualised as a CC lattice with extra sample points on the centre of each face of the cell. If the source stimulus is not band-limited or is sampled below the Nyquist rate, ensuring that aliasing is unavoidable, then the FCC lattice has been proven to reduce the aliasing artefacts when compared to other grid lattices (Künsch et al. 2005). Despite being more efficient, the BCC and FCC lattices require more complex indexing and reconstruction filters due to their non-separable nature.

Unstructured data sets on the other hand may have some arbitrary arrangement or in some cases no determinable structure with the values being arranged in an irregular fashion throughout the volume. In general, as expected, methods for rendering unstructured data sets are slower than the equivalent structured data sets (Williams & Uselton 1999, Lacroute & Levoy 1994).

2.3.4 Partial Volume Effects

Regardless of data acquisition method, when acquiring volume data the features of interest in the source stimulus can be small compared to the size of the voxels in the volume data. This means that each value in the volume data maybe an aggregate value of the different features within that area. This is called the partial volume effect (PVE) and is particularly evident at boundaries between two different features, for example between two different tissue-types. This can create pseudo-structures that have the same colour and opacity as an area of interest causing misleading images. Techniques to mitigate this include increasing the volume resolution, applying methods to estimate the fraction of each material in a voxel (Souza et al. 2005), as well as standardising data acquisition and reconstruction methods to be able to better compare scans performed on a particular device (Soret et al. 2007).

2.4 Light Participation in the Volume

DVR involves modelling the transport of light in the volume medium. To explain the modelling of light we treat the volume as being made up of particles that interact with each ray r of light travelling through the volume (Max 1995). Three properties are taken into account absorption, emission and scattering.

In the case of absorption each particle within the volume absorbs the light from each ray r to some degree. The amount of absorption depends upon the density of the particles, the position s along the ray r, $I(s)$ the intensity of light at the position s and $\tau(s)$ the *extinction coefficient* which is a function of the number of particles per unit area. The probability of occlusion and therefore the amount of light that is absorbed through the medium is given by the derivation (Max 1995):

$$\frac{dI}{ds} = -\tau(s)I(s) \tag{2.4.1}$$

By allowing particles in the volume to emit light, renderings such as glowing gases can be produced, when combined with absorption this produces the classic volume rendering integral. For emission each particle in the volume emits light with an intensity $C(s)$. With $\tau(s)$ as defined earlier the emittance model of light can be deduced by the derivation:

$$\frac{dI}{ds} = C(s)\tau(s) \tag{2.4.2}$$

Finally scattering can also be used to model the light through the volume. In this case the intensity of a ray r can be scattered by a particle, this is known as *out-scattering*. The intensity of the ray may also be added to, this is due to rays scattered from other particles increasing the amount of light that hits the particle. This is referred to as *in-scattering*. In most cases scattering is not used when volume rendering due to the complexity involved with modelling the interaction of all rays with each particle in the volume.

2.4.1 Emission-Absorption Model

By far the most common model used when volume rendering is the emission-absorption model that combines the two terms into the following derivation:

$$\frac{dI}{ds} = C(s)\tau(s) - \tau(s)I(s) \tag{2.4.3}$$

The extinction coefficient $\tau(s)$ is usually referred to as the opacity in volume rendering and is equivalent to $\alpha = 1 - T(s)$ where $T(s)$ is the transparency of the ray from 0 to the

position s and defines the rate of occlusion within the volume medium (Max 1995). The precise value of which is normally assigned by the use of a transfer function, see Section 2.8. To determine the transparency of a material between s_0 and s_1 the following integral is used:

$$T(s_0, s_1) = exp(-\int_{s_0}^{s_1} \tau(x))dx \tag{2.4.4}$$

This computes the absorption of the material, combining this with the emission section the full emission-absoprtion integral is produced:

$$I(D) = I_0 T(D) + \int_0^D C(s)\tau(s) \times T(s, D)ds \tag{2.4.5}$$

This equation integrates along the ray and models the absorption and emission of light from the position 0 towards the eye at position D. The terms I_0 refers to the intensity of the ray at the start of the volume (Max 1995) and $T(D)$ the transparency of the volume towards the eye position at D. This optical model is referred to as a *low albedo* approximation in that only a single scattering of light is allowed through the volume along the ray from the eye point e. This does limit the types of effect that can be produced, for example shadows require more than a single scattering of the ray. In the DVR literature this is referred to as the *volume rendering integral*.

2.4.2 Calculation of the Volume Rendering Integral

In most cases Equation 2.4.5 cannot be solved analytically, except when restrictions are placed on $C(s)$ and $\tau(s)$ (Max 1995, Moreland 2004). Therefore to solve the integral in the general sense an approximation must be made, commonly the method chosen is the Riemann sum. This can be used to approximate the value of an integral by splitting up the integral into n equal segments and successively summing the values of each segment. For the volume rendering integral the emission property can then be summarised as:

$$\int_0^D C(s)\tau(s) \approx \sum_{i=0}^{n} c_i \tag{2.4.6}$$

The absorption section of the integral can then be approximated as the product of the transparency of the n segments (Max 1995):

$$exp\Big(-\int_0^D \tau(x)dx\Big) \approx exp\Big(-\sum_{i=1}^{n} \tau(i\Delta x)\Delta x\Big) = \prod_{i=1}^{n} T_i \tag{2.4.7}$$

where $i\Delta(x)$ is a sample in the segment $(i-1)\Delta x \leq x_i \leq i\Delta x$. Finally then by combining the two approximations we have the full discretised volume rendering integral:

$$I(D) = \sum_{i=1}^{n} c_i \prod_{j=i+1}^{n} T_j \tag{2.4.8}$$

where c_i is the colour contribution of the ith interval and T_j is the transparency contribution of the jth interval.

Variations of the discrete volume rendering integral have been developed. This includes the semi-analytical method of Jung et al. (1998). Unlike the regular discrete approach, this calculates the colour and opacities between the n interval segments rather than at the sample points. This equates to a higher sample rate when using the regular fully discretised method.

2.5 Sampling Theory

Due to the simplification of the volume rendering integral and its approximation using the Reinman sum there is a choice to make regarding how many of the n equal sized samples must be taken to ensure that the volumetric data is accurately visualised.

This sampling problem is not unique to volume rendering, it is in fact part of the broader topic of signal processing. Given a continuous function $f(x)$ the goal is to create the most optimal discretised function representing the original signal without loss of information. This discrete signal is denoted as $f_s(x)$

Before proceeding to the sampling theory it is prudent to know just what is meant by sampling. In signal processing a sample is mathematically represented by a Dirac impulse function of the form in Equation 2.5.9 (Haddad & Parsons 1991):

$$\delta(x) = \begin{cases} \infty & \text{if x = 0} \\ 0 & \text{otherwise} \end{cases} \tag{2.5.9}$$

The reason such a function is useful is that by integrating the multiplication of the impulse function shifted by some value p with a function $f(x)$ the resulting value is the value of the function $f(x)$ at the position p.

In the spatial domain to create a discrete function $f_s(x)$ the original function $f(x)$ must be multiplied by a set of these impulse functions. This will provide the values of $f(x)$ at each of the positions of the impulses. The set of impulse functions is known as a comb and is defined in Equation 2.5.10. The comb function has a value of 0, except for

the intervals that correspond to the samples that have the value of 1:

$$comb(x) = \sum_{n=-\infty}^{\infty} \delta(x - nT_s) \tag{2.5.10}$$

where the value T_s controls the spacing between samples, in volume rendering this is equivalent to the distance between segments when performing the approximation of the Reinman sum. The continuous sampled function is therefore equivalent to:

$$f_s(x) = \sum_{n=-\infty}^{\infty} f(x)\delta(x - nT_s) \tag{2.5.11}$$

2.5.1 Nyquist and Shanon Theorem

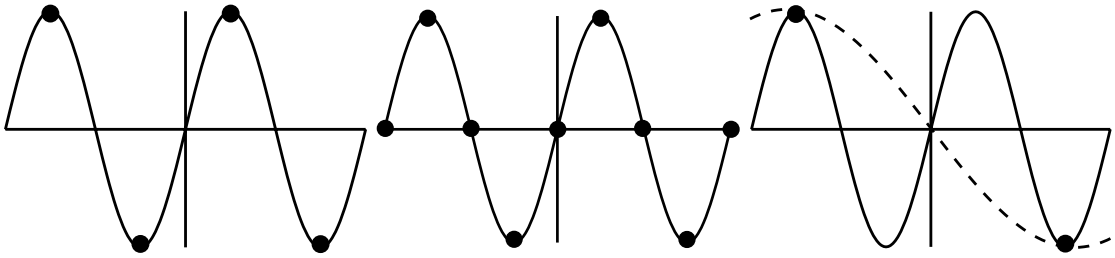
To determine the precise spacing of samples to accurately recreate a signal the Nyquist and Shannon theorem can be used (Shannon 1949). The theorem states that if a function is band-limited and contains no higher frequencies than f_{max} then it can be completely determined by a series of samples spaced no more than $\frac{1}{2f_{max}}$ apart. We define a function to be band-limited if the frequencies are contained entirely within the interval $[-f_{max}; f_{max}]$ where f_{max} is the bandwidth of the function. This implies that sampling below twice the highest frequency of the signal will induce *aliasing* where as sampling somewhat above this frequency will not increase the amount of information.

2.5.2 Aliasing

Sampling can be visually shown in the 1D case with a sine wave that has the highest frequency of f_{max} and the lowest frequency of $-f_{max}$. In this case given an ideal reconstruction filter, explained in the next section, by sampling at twice the frequency the sine wave in Figure 2.1a can be reconstructed. Sampling more than twice the highest frequency, and therefore above the Nyquist rate, also recreates the sine function 2.1b. Sampling at below the Nyquist limit however results in the dotted sine wave in Figure 2.1c, distinctly different from the original signal.

This type of aliasing is known in the literature as *pre-aliasing* or sometimes just as aliasing (Thévenaz et al. 2000*a*) and occurs if either the sampling rate is below the Nyquist limit or the data have not been suitably band-limited to remove high frequencies.

In terms of volume rendering this would suggest that in order to accurately sample the data, first the highest frequency of the measured stimuli must be gathered and then the sample rate should be set at twice this value. This poses a problem in that a large



(a) Sampling at Nyquist Rate (b) Sampling above Nyquist Rate (c) Sampling below Nyquist Rate

Figure 2.1: Sine wave function sampled at different rates

proportion of data for volume rendering is acquired from real-life stimuli, for example from medical scanning devices. This type of data does not lend itself to being band-limited. If the boundaries of the materials in the scene are sharp, arbitrarily large frequencies can be produced in the frequency domain voiding the sampling theorem. This means that any volume rendering of real-life data must necessarily produce some aliasing artefacts. Despite this, medical scanners in particular do not sample the real-life stimuli using a perfect dimensionless sample point. The input signal is actually the averaged value over a small region of space that has a point spread function similar to a Gaussian profile (Lacroute & Levoy 1994, Moller & Machiraju 1997, Sramek & Kaufman 1998). This averaging of values over a region acts as a low-pass filter and therefore removes high frequencies from the scene.

2.5.3 Frequency Domain

In the spatial or time domain a graph of a function maps the changes of a signal over space or time. The frequency domain however represents the signal as a range of frequencies.

Fourier's work states that any repeating function can be expressed in terms of the sum of sines and cosines at different frequencies with each signal being multiplied by some coefficient or amplitude. The offset of the sine waves from each other is called the phase shift. In the continuous setting for a function $f(t)$ the Fourier transform is denoted by $\mathscr{F}\{f(t)\}$ and the equation for defining the Fourier transform is given as:

$$\mathscr{F}\{f(t)\} = F(s) = \int_{-\infty}^{\infty} f(t)e^{-j2\pi st}dt \tag{2.5.12}$$

where $j = \sqrt{-1}$. An important property of this transform is that $f(t)$ and $\mathscr{F}\{f(t)\}$ are equivalent representations of the same function, in a practical sense this means that when a function has been transformed into the frequency domain it can be transformed back

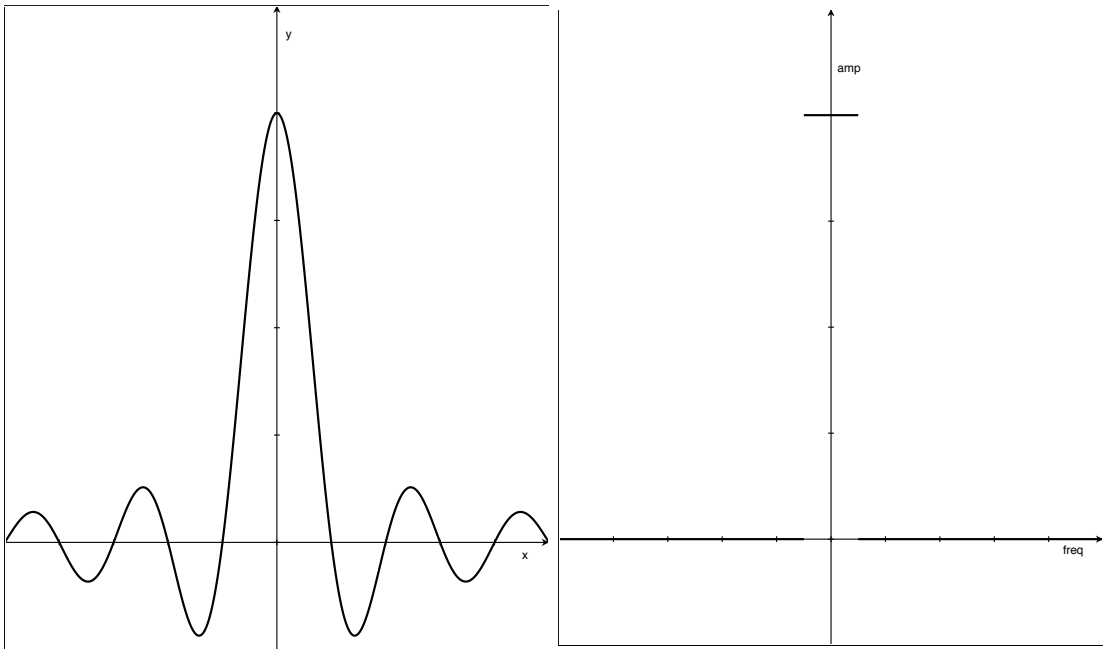
into the time or spatial domain without loss of information by using the inverse Fourier transform:

$$\mathscr{F}^{-1}\{F(s)\} = \int_{-\infty}^{\infty} F(s)e^{j2\pi st}ds \tag{2.5.13}$$

This transform however only applies to continuous functions, in the realm of image processing and volume rendering we often deal with finite length signals and so a modified form is required. This is known as the Discrete Fourier Transform (DFT), denoted by F_m and is given as:

$$F_m = \sum_{n=0}^{M-1} f_n e^{-j2\pi mn/M} \tag{2.5.14}$$

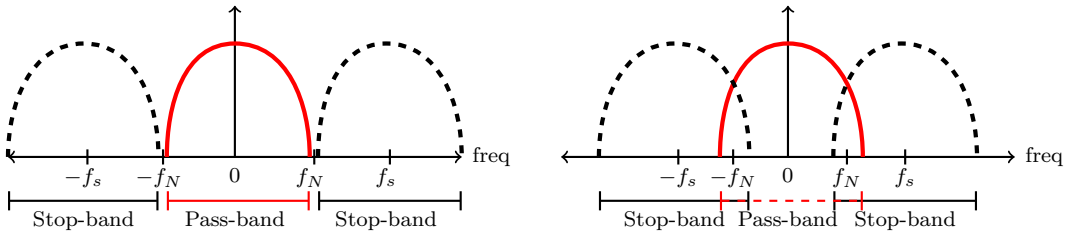
In this case the f_n corresponds to M set of samples from the original signal $f(t)$, and F_m then becomes the Fourier transform of the finite signal f_n. Figure 2.2 shows a *sinc* signal in the spatial domain and its transform in the frequency domain.



(a) Spatial response of *sinc* filter (b) Frequency response of *sinc* filter

Figure 2.2: Response of the *sinc* function in the spatial domain and frequency domain respectively. The spatial response has been clipped, in reality the lobes carry on infinitely.

The return value F_m is actually a complex number that encodes the amplitude and phase shift of the function in the frequency domain. In most cases when plotting the



(a) Adequate sampling in the frequency domain (b) Inadequate sampling in the frequency domain

Figure 2.3: Representation of sampling in the frequency domain

Fourier transform of a signal, magnitude is plotted against frequency with the phase shift of each sine wave being ignored (Foley et al. 1990).

2.5.4 Sampling in the Frequency Domain

In the frequency domain the process of sampling the continuous function duplicates the function to the positions of the impulses in the comb, shown in Figure 2.3a at a rate of 2π times the sampling frequency. The replication of the continuous periodic function centred at the origin is called the *pass-band* or *primary spectrum* and the copies of the function are called the *stop-band* or *alias spectra* (Marschner & Lobb 1994).

In the frequency domain, contrary to the spatial domain, for a high sampling rate the different spectra of the signal will be spaced further apart and for a low sampling rate the spectra will be spaced closer together. If the Nyquist rate is matched and the signal is band-limited then the replicas will be distinct, shown in Figure 2.3a. If however the Nyquist limit is not matched or the signal is not band-limited then the replicas will overlap as shown in Figure 2.3b.

In the process of interpolation, when viewed in the frequency domain, the job of the filter is to remove the replicas of the signal and pass the baseband undistorted (Mitchell & Netravali 1988). If the Nyquist rate is not observed or the function is not band-limited then the replicas will overlap and it will be impossible for any filter to correctly siphon out the original function.

2.6 Reconstruction

During the rendering process there is often the need to retrieve a value at the position (x, y, z) that does not lie on a a specific grid coordinate. The discrete data set must therefore be *interpolated* or *reconstructed* back into a continuous data set. Just as with

sampling, reconstruction is not limited to usage in volume rendering with the history returning to 1D signal processing and later having found much usage in image processing (Schreiber & Troxel 1985, Mitchell & Netravali 1988). Following the need to determine the accuracy of reconstruction filters for volume rendering, there is a relatively large body of work that can be called upon (Marschner & Lobb 1994, Moller & Machiraju 1997, Artner et al. 2005, Parrott & Stytz 1993, Theußl et al. 2000, Carlbom 1993).

In signal processing the process of reconstructing the continuous signal $f_r(x)$ from the sampled signal $f_s(x)$ involves *convolving* $f_s(x)$ with a kernel function $k(x)$:

$$f_r(x) = f_s(x) * k(x) \tag{2.6.15}$$

For a discrete signal $f_s(x)$ the convolution is described by the summation:

$$f_r(x) = \sum_{i=-m+1}^{m} f_s(i)k(u-i) \tag{2.6.16}$$

where $u = x - \lfloor x \rfloor$ and m is half the filter width. In practical terms to obtain a single point y in $f_r(x)$ the kernel function $k(x)$ is reflected about its origin and shifted by the amount i to the right. To produce the value at i the product of the f_s and k is created and then integrated, this is then repeated for all values of i. This equates to taking a weighted average of the neighbourhood around each point in the signal $f(x)$ where the weightings are provided by the flipped filter $k(x)$ producing the filter $f_r(x)$.

There is an important property of convolution regarding the relationship between convolution in the frequency domain and the spatial domain. The theorem states that the Fourier transform of the convolution of the functions $f(x)$ and $k(x)$ is equivalent to the product of the Fourier transforms of $f(x)$ and $k(x)$. Similarly given the product of the Fourier transforms of $f(x)$ and $k(x)$, by using the inverse Fourier transform the convolution in the spatial domain can be obtained. In volume rendering this relationship forms the basis of frequency domain rendering where interpolation can be performed by a wider filter in the frequency domain (Malzbender 1993, Dunne et al. 1990). This is possible because of the finite support of the *sinc* filter in the frequency domain.

Reconstruction filters can be classified into different groups, *separable*, *spherically symmetric* and other more exotic forms of interpolation functions. In 3D a filter $k(x, y, z)$ can be described as being separable if the filter can be written as the product of n filters, in volume rendering with a 3 dimensional data set a separable filter takes the form $k(x)k(y)k(z)$. Separable filters are more efficient in computational terms than other methods as they can

be performed in separate passes for each of the axis of the data set (Mitchell & Netravali 1988). As an example for a filter that has the width of N, a non-separable filter $k(x, y, z)$ can be performed in $O(N^3)$ where as a separable filter can be performed in $O(N)$. Instances of separable filters include the nearest-neighbour, linear and spline filters. A spherically symmetric filter or sometimes a *radially* symmetric has a value that only depends on the value of the filter from the origin. Examples of these filters include a rotated version of the cosine-bell as well as specific forms of the windowed sinc function (Marschner & Lobb 1994).

2.6.1 Ideal Reconstruction

In theory it is possible to perfectly reconstruct a continuous function from its sampled data set, providing that the continuous data source was band-limited before sampling. A normalised *sinc* function can be used in the 1D setting to perfectly reconstruct a continuous signal and is given as (Shannon 1949).

$$sinc(x) = \frac{sin(\pi x)}{\pi x} \tag{2.6.17}$$

For the 3D form the function is:

$$h_1(x, y, z) = (2f_N)^3 sinc(2f_N x) sinc(2f_N y) sinc(2f_N z) \tag{2.6.18}$$

where f_N denotes the Nyquist frequency of half the sampling rate along each axis (Marschner & Lobb 1994). This is also known as the rect filter due to its shape in the frequency spectrum. Figure 2.2 shows the spatial response and the frequency response of the sinc filter. In order to filter only the part of the frequency we are interested in the full width of the rectangle in the frequency domain should be equal to f_{max}, the bandwidth of the function. This can be shown by its definition in the frequency space as:

$$sinc(x) = \begin{cases} 1 & \text{if } |x| \leq \frac{f_{max}}{2} \\ 0 & \text{if } |x| > \frac{f_{max}}{2} \end{cases} \tag{2.6.19}$$

Whilst in theory the sinc filter can perfectly reproduce a signal, in practice it cannot be used due to it having infinite support in the spatial domain that requires all input samples to be used for the output of a single value (Engel et al. 2006, Bentum 1996) in technical terms this is a filter that has an Infinite Impulse Response (IIR) (Glassner 1995). When using frequency domain rendering it has however been shown that the sinc filter can be used to produce full volume renderings (Artner et al. 2005). This is due to the

frequency response of the sinc filter not having an infinite support in the frequency domain. Despite this impressive feat, it is still rarely used in practice due to the complexity of the approach as well as difficulties involved with performing the Fourier transform on GPUs if interactivity is required. Regardless of the use of the ideal filter, if the input signal is not sampled at the Nyquist frequency then in the frequency domain the replicas will overlap and no reconstruction filter will be able to recreate the exact function.

In general more practical reconstruction schemes are used to retrieve the off-grid values that have a Finite Impulse Response (FIR) (Glassner 1995). This must come with the knowledge that each method must therefore produce some level of errors and that the closer the reconstruction function comes to the frequency response of the sinc function, the more expensive the reconstruction function will be to calculate.

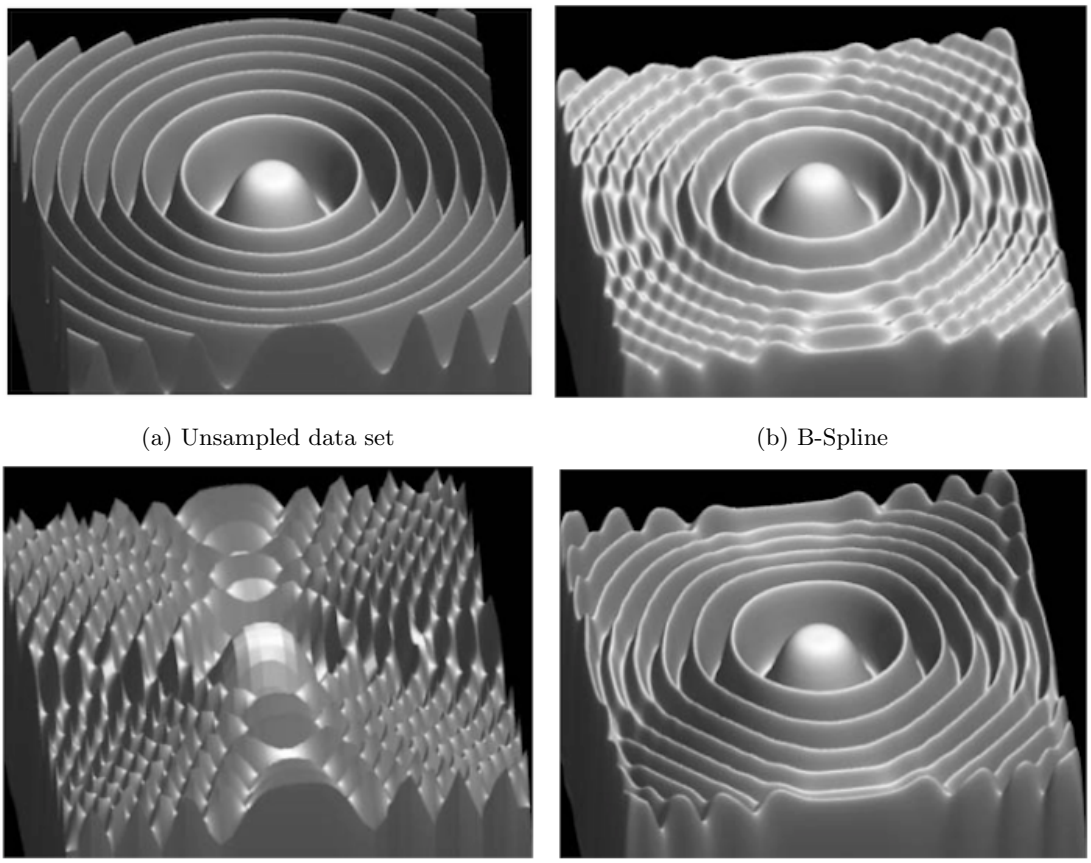
2.6.2 Reconstruction Errors

Practical reconstruction schemes must suffer from some form of error due to the non-ideal nature of the filter. The literature classifies the errors caused by reconstruction in a number of ways, either as *reconstruction errors*, or as general aliasing errors. The classification used here is that of Schreiber & Troxel (1985), Mitchell & Netravali (1988) and Marschner & Lobb (1994), where errors caused by under-sampling are referred to as aliasing errors and the other reconstruction errors are classified into post-aliasing, smoothing and ringing. In Figure 2.4 a high-energy sinusoid function, dubbed the Marschner and Lobb function has been recreated with multiple interpolation functions. It is apparent from the images that interpolation can have a significant impact on the quality of the rendering produced.

Post-Aliasing

Visually similar to aliasing caused by under-sampling of the data set, *post-aliasing* is different in that it refers to aliasing issues caused during the reconstruction stage of the DVR pipeline and particularly the use of a non-ideal reconstruction filter. In images post-aliasing can cause artefacts similar to those produced via under-sampling: jagged lines, staircase artefacts and spurious fine details.

The cause of post-aliasing is best described in the frequency domain. In this domain, the ideal reconstruction scheme has a spectrum of the box shape that abruptly cuts off at the bandwidth f_{max}. Ensuring that only the passband is covered. Non-ideal filters do not terminate at the f_{max} and hence the replications of the signal in the alias spectra in



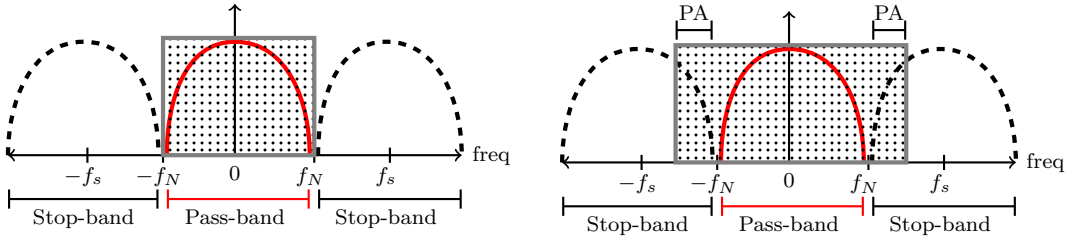
(a) Unsampled data set

(b) B-Spline

(c) Trilinear

(d) Windowed Sinc (cosine bell radius 4.7)

Figure 2.4: Four different images of the Marschner and Lobb data set. Fig (a) is the unsampled signal produced by ray-tracing. As can be seen Trilinear interpolation suffers aliasing issues producing the spikes in the image and ringing artefacts in the sides of the volume. The B-Spline function does not suffer from the more obvious aliasing issues but shows a considerable amount of smoothing. The windowed-sinc filter using a cosine bell with radius 4.7 shows the least amount of post-aliasing and smoothing issues although some amount of ringing is still present. These images were produced by Marschner & Lobb (1994)



(a) Ideal pass-band and stop-band behaviour

(b) Non-ideal stop-band behaviour

Figure 2.5: Visualisation of ideal and non-ideal stop-band behaviour leading to post-aliasing (PA).

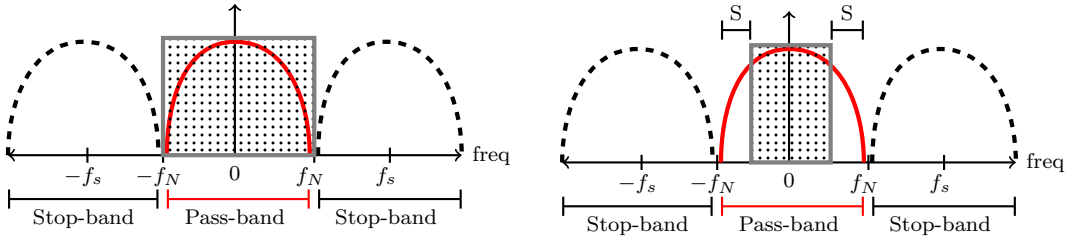
the stop band are not discarded and 'leak through' into the reconstruction (Marschner & Lobb 1994). The leaking of the stop band causes the aliasing artefacts. Figure 2.5a shows the behaviour of the ideal reconstruction scheme in the frequency domain, where as Figure 2.5b shows the behaviour of a non-ideal reconstruction scheme.

The sample rate of the reconstruction also has an impact on post-aliasing. In Figure 2.5a the spectra do not overlap but are very close together, this means that the sampling rate is very near the Nyquist frequency. If more samples were to be taken then the spectra would be placed further apart. The non-ideal interpolation scheme would then pass less of the alias spectra into the final rendering.

A particular form of post-aliasing is known as *sample-frequency-ripple*. This appears as oscillations in the reconstructed volume at the sample frequency. The cause of the ripples is due to a filter that is non-zero at the positions representing the lattice points in the frequency domain. Only filters that are non-zero at these points will not display any form of sample-frequency ripple. *Near-sample-frequency-ripple* is a subset of this problem that occurs due to non-zero filters near the lattice points of the grid in the frequency domain. In 2D images such ripples are visualised as patterns in the background that are not present in the original data source.

Smoothing

Smoothing is when rapid variations or high frequencies in the data are removed due to the use of a non-ideal reconstruction filter (Marschner & Lobb 1994, Bentum 1996). In images smoothing is represented as the blurring of fine details and the loss of sharp features. Compared to post-aliasing, smoothing occurs when the reconstruction scheme does not cover the entirety of the pass-band. Figure 2.6a shows the behaviour of the ideal



(a) Ideal pass-band and stop-band behaviour

(b) Non-ideal pass-band behaviour

Figure 2.6: Visualisation of ideal and non-ideal pass-band behaviour leading to smoothing (S).

reconstruction scheme in the frequency domain, where as Figure 2.6b shows the behaviour of a non-ideal reconstruction scheme that exhibits smoothing.

In some cases smoothing of the data can be a positive effect as it can reduce the amount of noise in a volume. However when there is too much smoothing, fine structures in the volume can be removed. Take for example the use of angiography imaging where fine blood vessels are displayed, smoothing can potentially remove these features. In the Shear Warp and Splatting volume rendering algorithms, smoothing can be a particular problem, in the former due to the number of sampling stages that can lead to the loss of sharp details (Lacroute & Levoy 1994) and in the latter due to the use of a Gaussian filter that can naturally blur the image (Novins 1993, Meißner et al. 2000).

Further smoothing can occur due to low-pass filtering of the original data. If the original data contains high frequencies that do not permit sampling at the Nyquist rate then a low-pass filter can be applied to attenuate the high frequencies. The removal of the high frequencies will smooth the data set. Although this is not technically a reconstruction issue, occurring before this stage, the impact on the visualisation of the data set is similar to that caused by non-ideal reconstruction filters.

Ringing

Ringing or *overshoot* is defined as the visual display of oscillations in the output signal near sharp transitions of values in the input signal. In images ringing is represented by ripples of high and low intensity pixels centred around sharp edges in the image. In most cases ringing should be avoided, however a single ring can help to sharpen an image (Mitchell & Netravali 1988). Any more than a single ring however produces negative effects. In signal processing terms this is known as the Gibbs phenomenon with two primary causes, the

use of a low-pass filter or by using a truncated sinc reconstruction filter.

For a band-limited signal the low-pass filter, that passes signals below the Nyquist frequency, can be used to perfectly reconstruct the signal. Natural signals, for example body parts, are rarely band-limited and so in this case the low-pass filter can be used to band-limit the input signal prior to the sampling and reconstruction stage. If the low-pass filter is used prior to sampling the sharp cut-off of the filter will cause ringing artefacts around sharp discontinuities in the input signal. These ringing artefacts will persist regardless of the choice of the filter used in the later reconstruction stage.

To deal with the problem the truncated sinc filter can be used as a practical alternative to the ideal sinc filter. The truncated sinc filter in the frequency domain does not have the perfect rectangular shape of the sinc filter and therefore frequencies that are suppressed in the standard sinc filter are passed and it also attenuates and amplifies other frequencies around sharp transitions. By increasing the width of the truncation and therefore being closer to the true sinc filter the problem can be lessened although the actual ringing will occur as a long as the filter has finite support. The ringing artefacts are most perceptible around edges of high contrast where the peaks and troughs of the ringing will be greatest.

A visually similar problem to ringing is caused by *negative side lobes*. This is where the response of the filter in the spatial domain dips below the x-axis. By convolving a signal with a filter that contains negative side lobes, negative intensities can be created in the reconstruction. If for example these intensities correspond to pixel colours then the negative values will need to be clamped.

Anisotropy

Non-spherically symmetric filters cause a condition known as *anisotropy*, this results in asymmetric artefacts in smoothing and post-aliasing (Marschner & Lobb 1994). The amount of post-aliasing and ringing will differ depending on the orientation of of features contained within the volume in relation to the orientation of the filter.

2.6.3 Practical Reconstruction Filters

A selection of practical reconstruction filters will be detailed. In each of the descriptions the *regularity* of the function will be given. The regularity of the function is the N number of times that a continuous function can be differentiated. A discontinuous function cannot be differentiated and has no regularity. A function that is continuous but not differentiable

will have a regularity of C^0 (Thévenaz et al. 2000*b*).

Nearest-Neighbour

The simplest and fastest method of reconstruction is the nearest-neighbour filter, also known as the sample-and-hold filter or the box filter. Whilst arguably not a true filter, given a sample point (x, y, z) the point is simply rounded to the nearest grid value. As the function is not continuous it has no regularity.

$$NearestNeighbour(x) = \begin{cases} 1 & \text{if } |x| < 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{2.6.20}$$

Poor image quality is returned when using this method unless the sampling and coordinate systems have the same spacing or when the data set is constant (Bentum 1996).

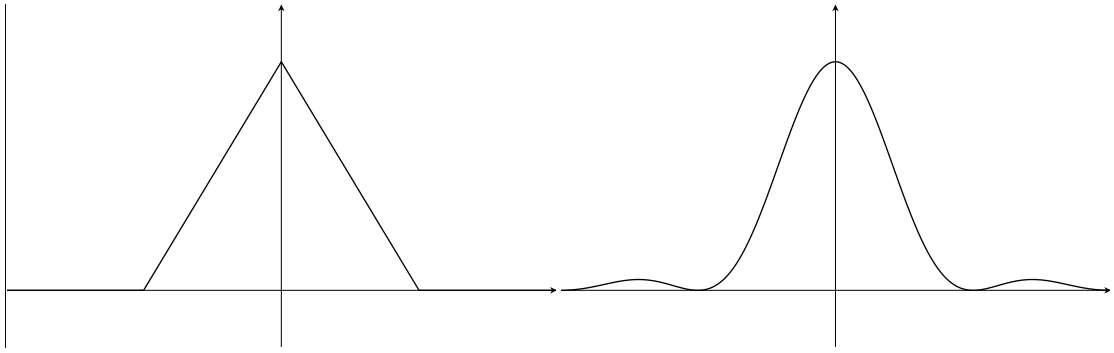
Linear

By far the most common, due to its prevalence in GPUs, speed and adequate image quality is the trilinear reconstruction scheme. As the name might suggest this is a combination of linear reconstructions along the three axis that generates a value based on the eight coordinates of the grid surrounding the requested point. The linear filter is continuous but cannot be differentiated and so has a regularity of C^0. In the 1 dimensional form the reconstruction scheme is defined as:

$$Linear(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.6.21}$$

Due to the filter assuming linear interpolation, high frequency data or data that does not vary linearly along the three axis of the volume will not be captured correctly in the output image (Marschner & Lobb 1994). Under very specific circumstances when the interpolation scheme is tailored to the dataset tolerances can be set on the quality of the interpolation (Zheng et al. 2010). However in general the limited number of neighbourhood voxels used sets a cap on the capability of the filter, this is because the eight voxel values used in the vicinity of the sample point are not enough to adequately describe the variability of the data set (Parrott & Stytz 1993).

A form of modified linear interpolation has been proposed in the paper Blu et al. (2004) that shifts the sample points for the linear interpolation. In experiments with 2D image rotation the quality of the modified filter has been found to be close to that of cubic interpolation with a significantly reduced complexity.



(a) Spatial response (b) Frequency response

Figure 2.7: Fig (a) shows the spatial response of the linear filter, note the finite width of the filter. Fig (b) shows the frequency response of the linear filter.

The spatial response of the 1D case is the tent filter in Figure 2.7a. In the frequency domain the spectrum of the filter is described by the $sinc^2$ function, as can be seen in Figure 2.7b.

Splines

A spline is a piecewise polynomial function that is defined by a finite set of connected pieces. A specific form of these splines known as B-splines have been used in image interpolation (Andrews 1978, Mitchell & Netravali 1988) and have been adapted for use in volume rendering (Levoy 1987). The closed form of a B-Spline, defined as B^n where n is a natural number ≥ 0 is (Thévenaz et al. 2000a):

$$B^n(x) = \sum_{k=0}^{n+1} \frac{(-1)^k(n+1)}{(n+1-k)!k!}(\frac{n+1}{2} + x - k)_+^n \tag{2.6.22}$$

where $(x)_+^n = (max(0,x))^n$. Although B-splines can be made to any arbitrary degree, a spline of B^0 is identical to the basic nearest neighbour reconstruction up to a translation of the origin. A spline of B^1 is equivalent to the convolution of $B^0 * B^0$, or the convolution of two box functions that produces the tent function, in other words the linear interpolation scheme.

The B-Spline with $n = 3$ is known as the cubic B-spline. The definition of the basis function of the cubic spline is:

$$B^3(x) = \begin{cases} \frac{2}{3} - \frac{1}{2}|x|^2(2-|x|) & 0 \leq |x| < 1 \\ \frac{1}{6}(2-|x|)^3 & 1 \leq |x| < 2 \\ 0 & 2 \leq |x| \end{cases} \tag{2.6.23}$$

As a B-spline is a piecewise polynomial with degree n, it has a regularity of C^{n-1}. An advantage of using such splines is that the frequency spectra of such a function is strictly positive. A modification of the B-spline is presented by Keys (1981), that does not require a modification stage so that higher order filters are made to be interpolating. The study of splines is further extended to generate the BC-Spline category of functions (Mitchell & Netravali 1988). Specifically designed to remove the problem of sample-frequency ripple, a form of aliasing occurring at sample points in the frequency domain, the filter is adapted so that the frequency response of the function is zero at the integer multiples of the sampling frequency except at 0. That means that only two parameters are required, B and C. All interpolation functions from this family are C^1 continuous, meaning that the first derivative of the function can be found. From subjective evaluations performed by Mitchell & Netravali (1988) it was found that functions that lie on the line $2C + B = 1$ produce satisfactory results. In comparison from evaluations in the frequency domain Marschner & Lobb (1994) found no particular reason to prefer filters over the $2C + B = 1$ line due to a need to balance post-aliasing and smoothing artefacts.

The cubic and later BC-splines use a total of four grid points in order to calculate the reconstructed value. A spline that uses two points can also be constructed, such a function is known as a 2-point spline (Bentum 1996). As the spline uses only two-points it shares certain similarities with linear interpolation that likewise uses the same number of points in order to reconstruct a value. A particular problem noted by Bentum is that this spline is not capable of reconstructing straight lines unless the sample rate approaches zero. When testing with analytical renderings of a high frequency function in a 2D image there was found to be little difference between linear and 2-point splines.

Although classically splines are more computationally complex than other interpolation schemes, efficient implementations for GPUs have been developed that require only eight trilinear texture fetches in order to evaluate the cubic B-spline for volumetric data (Sigg & Hadwiger 2005). Building upon this work a volume rendering scheme is developed that uses the weighted average of two B-Spline interpolation schemes, a standard B-spline and a pre-filtered B-spline interpolation with the weighting being controlled by the parameter r (Csébfalvi & Domonkos 2009). By allowing a user to interactively modify the parameter the trade-off between the post-aliasing and smoothing errors can be controlled.

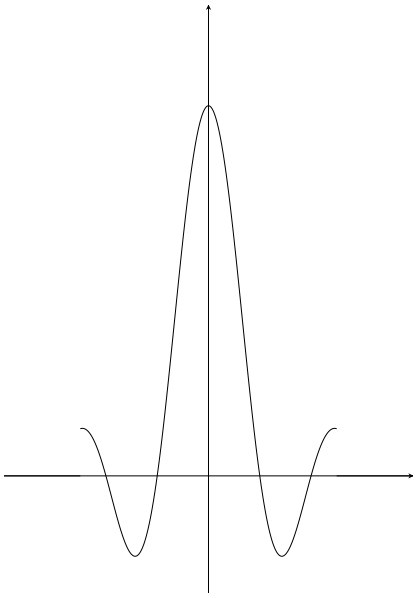


Figure 2.8: Spatial response of the truncated sinc function

Modified Sinc Functions

Whilst the pure sinc function cannot be used directly due to its infinite support, modified forms of this function can be. The goal of such a filter is to approximate the ideal sinc filter to enable practical reconstruction whilst still keeping most of the desirable properties. The approximation is made by narrowing the support of the sinc function so that it no longer has infinite support in the spatial domain. This does mean that the response of the filter in the frequency domain will no longer be ideal and so some reconstruction errors must be present. A method of reducing the complexity of the sinc function is by multiplying the sinc function with some window function that is zero outside of some radius r.

The simplest method is by simply truncating the width of the sinc function. This is equivalent to multiplying the sinc filter with a box filter of the form:

$$\Pi_r(x) = \begin{cases} 1 & \text{if } |x| \leq \tau \\ 0 & \text{otherwise} \end{cases} \tag{2.6.24}$$

where τ is the radius of the window function. The 1D truncated form of the sinc function is shown in Figure 2.8, note how the waves cut off abruptly. The clipping of the sinc function causes ringing, a reconstruction error whereby oscillations in the output appear near high frequency edges. Prior studies have shown that the quality of images produced when using the truncated sinc filter is low and in some cases not considerably better than the Nearest-Neighbour interpolation scheme (Bentum 1996).

In order to prevent the severe ringing artefacts that are present when simply truncating the sinc function a different approach must be used. By multiplying the ideal reconstruction filter with a function that smoothly decreases to 0 the most severe forms of aliasing are alleviated. The filter will still be 0 outside some radius r, however the waves reduce smoothly.

There is a large number of possible window functions that can be used and each comes with a different computational cost and variable quality that is unique to that window. One thing each of the windows share in common is that the radius of the window can be made arbitrarily large. The result of this is that the windows with large radius will produce fewer artefacts in the reconstruction. The obvious issue is that the larger the radius and the closer it is to the sinc function then the more expensive it will be to evaluate, likewise the smaller the radius then the less similar the function will be to the ideal filter.

Examples of windowing functions that can be used are the cosine bell, the Bartlett (a tent function), Welch, Parzen, Blackman and Kaiser. In testing different reconstruction issues with volume rendering, Marschner & Lobb (1994) found that a cosine bell window with a radius of 4.78 produced fewer amounts of smoothing and post-aliasing artefacts than any cubic spline filter. However the filter caused a certain amount of anisotropy which led to variations in height of the reconstructed volume. In a review of different window functions Theußl et al. (2000) found that the optimum radius of a window function was 2, with increasing radius size not bringing about a significant increase in the quality of the reconstructed volumes. From further evaluations based on the Taylor series as described by Moller & Machiraju (1997) it was determined that the Blackman, Kaiser and truncated Gaussian window produced fewer errors. This was further shown by high quality reconstructions of the Marschner and Lobb test data set. The definitions of such filters are as follows:

$$Gaussian(x,\tau,\sigma) = \begin{cases} 2^{-(\frac{x}{\sigma})^2} & \text{if } |x| < \tau \\ 0 & \text{otherwise} \end{cases} \tag{2.6.25}$$

$$Kaiser(x,\tau,\sigma) = \begin{cases} \frac{I_0(\alpha\sqrt{1-(\frac{x}{\tau})^2})}{I_0(\alpha)} & \text{if } |x| < \tau \\ 0 & \text{otherwise} \end{cases} \tag{2.6.26}$$

In regards to these functions there are new parameters that can be seen, the σ variable in the Gaussian window is the standard deviation of the function. This is used to control the support of the filter to produce a computationally viable filter. The I_0 is the zeroth order Bessel function and the α is similar to the σ function in the Gaussian filter, defining

the inverse width of the Kaiser window. The regularity of a windowed sinc is not straightforward and depends upon the windowing function that is used to narrow the width of the sinc function.

Pass-band Optimal

A family of interpolation filters introduced to volume rendering by (Carlbom 1993) are referred to as *pass-band* optimal filters. A pass-band optimal filter is defined as a linear combination of discrete filters. Each of which uses weighted Chebyshev minimization to obtain a filter that has a Fourier transform with a minimum distance from the ideal frequency up to some frequency $f_m < f_N$ where f_N is defined as the Nyquist frequency of the data set. The resulting filter minimises the amount of smoothing that is exhibited in the final reconstruction.

An alternative form of pass-band optimal filters are introduced by (Csébfalvi 2008). In this case the discrete Fourier transform of the volume is pre-multiplied by the reciprocal of a standard reconstruction filter. The volume is then converted back into the spatial domain and reconstructed using the same filter. This method optimises the pass-band behaviour of the chosen filter. In both cases although smoothing is significantly reduced, such filters may have arbitrarily bad behaviour in the stop-band and therefore significant post-aliasing.

2.6.4 Comparisons of Reconstruction Filters

The choice of reconstruction filter to use is one of the most important algorithmic choices in DVR, alongside the sampling rate and transfer function used to classify the data. There exists different ways in which reconstruction filters have been tested, below evaluations of reconstruction filters are classified according to the approach used to compare them.

Subjective Evaluations

Perhaps surprisingly, given that reconstruction methods have a direct visual impact on DVR images and are often used in 2D applications where the images will eventually be viewed by a person, there are few user-based studies of reconstruction schemes for DVR or even on reconstruction filters for 2D image processing. It is likely that there is a preference towards objective evaluation methods due to the signal processing background of the field,

where various properties of the reconstruction schemes can be evaluated without requiring user intervention or opinion.

Initial evaluations of reconstruction in 2D image processing were undertaken by Schreiber & Troxel (1985). Describing the spatial sensitivity of the visual system it was found that in some cases high quality reconstruction filters are not necessary, depending on the distance from the user to the display. In images where the spacing of each sample subtends an angle of 0.5 degrees the structure of the samples becomes highly noticeable. For sample spacing that is less than 0.1 degrees of visual angle then the sample structure is effectively averaged out and so becomes less noticeable. From this, the suggestion is that different reconstruction schemes are used at different distances of the user to the display and that at high sampling rates high-quality reconstruction schemes are not completely necessary. In experimental studies comparing a Gaussian filter as a windowing function with other reconstruction schemes including linear and nearest-neighbour, using photographic and text-based images, it was found that participants preferred the quality of images produced when a Gaussian filter was used. Regarding other filters, nearest-neighbour consistently produced low results with truncated sinc producing results in some cases only marginally better than nearest-neighbour.

The amount of blur, anisotropy and ringing have been tested with a class of spline functions called BC-Splines (Mitchell & Netravali 1988). In trials the two parameters of the function, B and C, were randomly set for a large group of images, participants were required to state which of either ringing, blurring or anisotropy the image exemplified. Settings with parameters B and C $= 1/3$ were found to provide an acceptably high level of image quality. A limitation of the experiments is that the manner of subjective testing used allowed only a single image quality matching to be made for each image. This prevents the ability to determine whether a certain set of parameters contained a high-degree of ringing and blurring combined.

Metric Based Evaluations

As well as the subjective methods for comparing different interpolation functions there are methods for comparing the quality of reconstruction schemes using some metric.

Marschner & Lobb (1994) define a set of metrics to asses the amount of smoothing, post-aliasing and ringing that each reconstruction filter exhibits by computing it in the frequency domain. Smoothing is assessed by measuring the energy of the reconstruction

filter in the pass-band region where an ideal filter will have the value of 1. The metric is defined as:

$$S(h) = 1 - \frac{1}{|R_N|} \int_{R_N} |\hat{h}|^2 dV \tag{2.6.27}$$

where h is the reconstruction filter to be measured, $\hat{h}$ is the frequency domain representation of h, R_N is the Nyquist region, $|R_N|$ is the volume of R_N in the frequency domain and dV is an infinitesimal element in the volume R. Post-aliasing is assessed by measuring the energy in the stop-band where a value of 0 indicates no post-aliasing errors. The metric is defined as:

$$P(h) = \frac{1}{|R_N|} \int_{\overline{R_N}} |\hat{h}|^2 dV \tag{2.6.28}$$

where $\overline{R_N}$ is the complement of the region R_N. Finally the ringing metric assesses the error when a specific reconstruction filter is convolved with a unit step function and is defined as:

$$O(h) = max(p_s * h) - 1 \tag{2.6.29}$$

where p_s is the unit step function. These metrics are used in Chapter 5, Chapter 6 and Chapter 7 to measure the properties of reconstruction filters when they have been used to render DVR images.

Although these metrics give the amount of smoothing, post-aliasing or ringing that a reconstruction filter exhibits they do not measure the characteristics a specific data set has on the artefacts. Instead, these metrics use only the frequency response of the filter. Therefore, these metrics may not be effect when performing comparisons across different volume data sets.

In contrast to the frequency domain metrics, properties of each reconstruction filter can also be measured in the spatial domain. The errors can be defined as the truncation error caused by a reconstruction filter with a narrow width and non-sinc error caused by using a filter other than the sinc, (Machiraju et al. 1995). The truncation error is affected by the distance between data points in the volume, the position of a chosen sample point and the selected reconstruction filter. Low error values are found when either the reconstruction filter used has a large width or the sample point is close to an existing data point in the volume. The non-sinc error measures the difference in areas between a chosen reconstruction filter and the sinc filter in a neighbourhood around a sample position. The closer the reconstruction filter is to the sinc filter, the smaller the error will be.

Keys (1981) and Moller & Machiraju (1997) use the Taylor series expansion where

different reconstruction filters are measured by comparing the remainder term when approximating the Taylor series to the Nth degree. Reconstruction filters can be further classified into categories depending on N, in this case filters that are in the class $N-1$ are defined as being Nth degree error filters (N-EF), where an N-EF filter can reconstruct a polynomial of $N-1$th degree or lower without errors. In most cases a filter with a higher degree will be more accurate than a filter with a lower degree unless low sampling rates are used.

2.7 Compositing

By using a process known as compositing RGBa colours can be layered (Porter & Duff 1984) allowing the accumulation of multiple semi-transparent surfaces. In direct volume rendering this allows the discretised volume rendering integral in Equation 2.4.8 to be processed serially by compositing each iteration of sampling to produce a final pixel colour for each ray. This produces the recursive *front-to-back* compositing scheme:

$$\begin{aligned} C_{dst} &= C_{dst} + (1 - a_{dst})C_{src}, \\ a_{dst} &= a_{dst} + (1 - a_{dst})a_{src} \end{aligned} \tag{2.7.30}$$

where C_{dst} stores the accumulated associated colour, C_{src} is the colour intensity of the current interval and a is opacity, an alternative form of the transparency T, where $a = 1 - T$. A benefit of the front-to-back compositing scheme is that once a_{dst} is equal to 1.0 at some layer i the algorithm can stop as light intensity from layers $i+1$ and beyond will be fully occluded (Kaufman & Mueller 2005).

The *back-to-front* composition scheme is an alternative to the above method where the layers are accumulated from the furthest distance towards the eye point. In this case the recursive scheme is:

$$C_{dst} = (1 - a_{src})C_{dst} + C_{src} \tag{2.7.31}$$

This method is often used in the texture mapping implementation of the DVR algorithm to blend the texture slices of the volume to generate the final image.

2.8 Classification

In the DVR literature classification is the process of applying optical properties, colour and alpha, to the scalar values of the volumetric data set, this is used to display structures and

surfaces within the volume. The function used to assign the colour properties is called the *transfer function*. This is commonly implemented as a 1D or 2D lookup table where the scalar value at the position (x_i, y_i, z_i) is used as a key for the lookup table and an RGBa colour value is returned. When using 2D transfer functions, the gradient at a sample point can be used as the second key, allowing surfaces to be represented more clearly.

2.8.1 Automatic Classification

By allowing users to interactively modify the transfer function different parts of the structure of the volume can be displayed without requiring the application to be recompiled. However, the creation of a suitable transfer function can be a non-trivial problem and can often be put down to a trial-and-error approach (Schroeder & Martin 2001). Therefore, due to the difficulty in creating suitable transfer functions, studies have taken place on the creation of automatic transfer functions. Techniques for generating transfer functions can be split into *data-centric* methods that use properties of the volume data to generate the transfer functions or *image-centric* methods that use information from the rendered image to modify the transfer function.

Initial data-centric approaches used the scalar values and derivatives to generate a histogram of boundaries in the volume which was then used to generate a suitable transfer function (Kindlmann & Durkin 1998). This method was later extended to generate a range of transfer functions with thumbnails of the rendered images being presented to the user to create a semi-automatic generation method (Prauchner et al. 2005). Neural networks have also been applied to data-centric transfer function generation. Tzeng et al. (2005) gather a painted region of the volume data by the user and use this as a training set for a neural network to classify the rest of the volume. In Zhang & Sun (2003) a general reguression neural network is used in the design process of a multi-dimensional transfer function.

Image-centric techniques typically require input from the user in order to provide information regarding the suitability of a transfer function after a DVR image is produced. A genetic algorithm can be used to produce a series of transfer functions where the fitness value of the transfer function can be manually given by the user after images are produced (He et al. 1996). Design galleries are another popular approach to assisted transfer function generation (Marks et al. 1997). A user interface is presented that contains a broad selection of preview images that can be produced by varying the transfer function param-

eter. Extensions to this method have been devised where the preview images displayed contain a high amount of information according to an entropy metric (Schlegel & Pajarola 2013).

2.8.2 Pre-classified versus Post-classified

The transfer function can be applied before or after the volumetric data has been interpolated using a reconstruction filter, this is called *pre-classified* and *post-classified* respectively. In the pre-classified state the scalar values in the volume are assigned colour and opacities, these colours and opacities are then interpolated to give values between sample points in the volume. In the post-classified method the scalar values in the volume are first interpolated and then these interpolated scalar values are assigned colours and opacities from the transfer function.

Pre-classified transfer functions can cause artefacts due to misrepresentation of high frequencies in the transfer function where as post-classified transfer functions are better able to capture theses high frequencies and therefore produce a rendering with higher image quality (Engel et al. 2006, Kaufman & Mueller 2005). This is primarily because pre-classification only accounts for the high frequencies in the volume, by classifying and then interpolating high frequencies introduced in the transfer function are ignored. Post-classification is able to account for high frequencies in the transfer function that might be introduced by the use of step functions.

An advantage of pre-classification is that when used with graphics hardware the volume data can be converted into a GPU texture containing the RGBa colours from the transfer function prior to rendering. This can decrease the rendering time required due to there being no need to request the colour from the transfer function for each interpolated point (Engel & Ertl 2002). In an interactive setting this benefit is negated by the need to reload the entire volume texture if the transfer function is changed.

When non-linear transfer functions are used, even post-classification can still suffer from aliasing artefacts caused by high frequencies. One solution is to increase the sample rate, however this would need to be recalculated each the time transfer function were to be changed. A second solution is to use what is called *pre-integrated* classification (Engel & Ertl 2002). This splits the calculation of the discrete rendering integral into two parts, the integration of the scalar values and then separately the integration of the transfer function. Colour and opacity of the transfer function are calculated by integrating between slices of

the volume rather than at the sample points. This has the benefit of reducing the number of samples needed as well as capturing high frequencies created by non-linear transfer functions, however colour and opacity between each sample must be recalculated if the transfer function is changed.

2.9 Gradient Estimation

When classifying the samples along the ray, it can be useful to know if a sample point belongs to a boundary of an object in the the volume. This is particularly useful when using a shading method such as Phong shading to add more realistic lighting to a surface in the volume. The gradients at sample points can also be used as a second key for a 2D transfer function to map different surfaces in the volume to separate opacities and colours.

This poses a problem as volumetric data does not contain representations of surfaces, being a set of grid points with some associated data. In surface rendering the normal vectors of the surface would be used for the calculation of shading, in volume rendering the normal is replaced by the gradient vector. The gradient vector then indicates the rate of change at the surrounding area.

As with the interpolation schemes, there are different methods for calculating the gradient vector from the data set. In the ideal sense the so called *cosc* filter can be used (Theul 1999*b*, Bentum 1996). This is the derivative of the sinc filter used for interpolating points and is given as:

$$cosc = \frac{cos(\pi x) - sinc(\pi x)}{x} \tag{2.9.32}$$

As it is the derivative of the sinc filter which has infinite support it cannot be computed efficiently and so simpler methods must be used. The most common method is the *central difference* method that computes an average of the difference along each axes. The gradient at a point (x, y, z) is given by the equation (Theul 1999*a*):

$$\begin{aligned} g_x &= \frac{f_{x+1,y,z} - f_{x-1,y,z}}{2} \\ g_y &= \frac{f_{x,y+1,z} - f_{x,y-1,z}}{2} \\ g_z &= \frac{f_{x,y,z+1} - f_{x,y,z-1}}{2} \end{aligned} \tag{2.9.33}$$

A similar method is the so called *intermediate difference* that is used to compute the gradient for sample points that lie between voxels. In this case an interpolation method is required to produce an estimate of the value between the original two voxels.

A somewhat better gradient estimator is the Sobel operator that uses 26 neighbouring points to produce an estimate. This comes with the expected increase in computation time and the increase in the use of the neighbouring cells can cause problems with smoothing (Bentum 1996). Other less common methods for gradient estimation include using cubic splines as well as using windowed versions of the cosc filter (Bentum 1996).

Comparisons of Gradient Estimation Schemes

The central difference method, despite being simple to compute, is susceptible to degradation caused by noise in data sets from ultrasound and MRI systems (Ropinski et al. 2010). Noise is a particular problem for gradients that are within a volume, these should naturally have no magnitude, being part of a surface and not in the intermediary regions between changing of surfaces, however noise within the volume can falsely produce gradients with some magnitude indicating false surfaces.

The Sobel, cubic and modified cosc derivatives, despite producing higher quality gradients, require a much greater percentage of processing time to compute.

In regards to comparisons between the more complex cubic and windowed truncated cosc filters there are differing opinions. Theul (1999*b*) conclude that the modified cosc filters produce images of better quality when simulated data sets are compared visually.

In certain cases the estimation of the gradient scheme can have a larger impact on the quality of the image than the choice of interpolation scheme used (Möller et al. 1996). Möller et al. (1996) devised a number of methods to assess the quality of gradient or derivative schemes based on analysis of Taylor series. However, we note that the use of a gradient scheme for assisting in classifying or shading the volume is optional, compared to the interpolation and sampling stage of the volume rendering pipeline that cannot be avoided.

2.10 Practical Implementations

Several techniques have been developed for rendering volumetric data sets, we will briefly review the most common algorithms in use. We will only concern ourselves with DVR algorithms. Where it is applicable we will describe some of the modification to the algorithms as well as artefacts that are native to each method. Each of the direct volume rendering techniques discussed have their own set of parameters that can affect image quality and computational cost and we will see that not all of the rendering algorithms

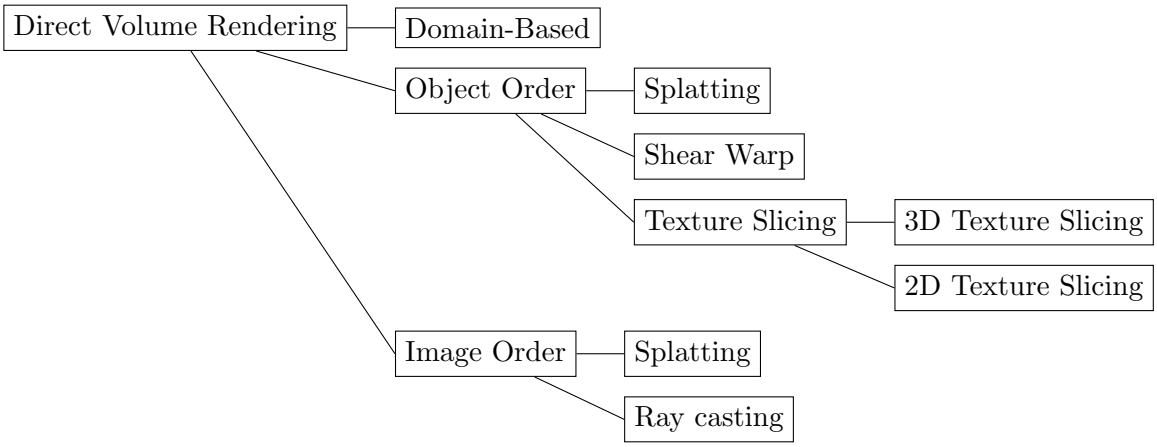


Figure 2.9: Taxonomy of DVR algorithms

are designed with the same goals.

There are three main types of volume rendering algorithms and these can be categorised by how they iterate when producing the rendering, they are: *image-order*, *object-order* and *domain-based*. Image-order algorithms iterate over pixels in the image. Object-order algorithms iterate over voxels in the volumetric data and can therefore make use of the spatial structure of the data set. Finally domain-based algorithms transform the data set into another domain, typically the frequency domain, and perform the rendering entirely in that domain. Figure 2.9 shows the taxonomy of direct volume rendering algorithms.

Shear-Warp

Popularised by Lacroute and Levoy (Lacroute & Levoy 1994) this is a hybrid approach between image-order algorithms and object-order algorithms and is the fastest of the software-based volume rendering methods (Meißner et al. 2000). It is designed to combine the efficient access of the data from object order approaches with the high image quality of the image-order techniques.

The volume is treated as a set of slices with the algorithm iterating through each slice of the data set in order. The projection takes place on an intermediate buffer known as the base plane. The slices within the volume are 'sheared' so that the base plane lies parallel to the slices in the volume. Rays are then traversed from this base plane through the volume. After projection to the base plane, 2D image processing techniques are applied to warp the base plane geometry of the final image plane.

Due to the intermediate base plane being parallel to slices of the volume, each line of pixels in the intermediate image is parallel to each line of voxels in the volume data. This

allows a run-length encoding method to be used to skip voxels that are fully transparent and therefore speed up the rendering process and decrease the memory requirement. Moreover, as the volume is divided into slices a 2D interpolation scheme is used instead of a more intensive 3D interpolation scheme further reducing the rendering time required, although this is at the cost of image quality.

Despite the cost that can be saved by the run-length encoding of the data set, when interactive rendering is used, three data sets are required, one for each of the three axis. The time and memory saved by the encoding scheme is therefore balanced by the multiple data sets used (Kaufman & Mueller 2005).

The sample rate of the algorithm is directly tied to the number of slices in the volume. When the number of slices is small aliasing is likely to occur. To increase the number of samples intermediary slices are required that slot in-between each of the existing volume slices and some method of interpolating between the intermediary slices and the existing volume slices must be used (Lacroute & Levoy 1994).

2.10.1 Splatting

Introduced by Westover (1990), this method is an object-order algorithm. Each voxel within the volume is projected onto the image plane after being assigned a colour and opacity based on the transfer function. The size and intensity of the footprint, or *splat*, of the voxel on the image plane is defined by a radially symmetric interpolation kernel, commonly a Gaussian function (Mueller et al. 1999). The amalgamation of each of the splats in a back to front order generates the final image.

In terms of rendering time, only voxels that will be present in the final rendering of the image are projected, a method was also proposed whereby splats that will be invisible are removed from the rendering pipeline (Meißner et al. 2000).

Due to the nature of the smoothly decaying radially symmetric splat, the images are naturally anti-aliased. Excessive blurring however can impact the image quality of the final rendering. Certain methods can be used to remove the blurring that include performing gradient and and shading calculations after the projection of the voxels and modifying the splats near edges so that they decay faster, causing less blurring around the edges (Mueller et al. 1999).

2.10.2 Texture Based Volume Rendering

Object-order techniques were adapted to run on graphics hardware in response to programmable GPUs capable of handling the large amounts of memory required for volume rendering. Using this approach the volume data is stored within the GPU as either a single 3D texture or a set of 2D textures depending on the capability of the hardware. In both cases the textures are applied to a set of flat layered polygons. The textured polygons are then rendered parallel to the viewing direction and blended back to front prior to being displayed. The number of polygon layers used to represent the volume depends on the number required to satisfy the Nyquist theorem. In practice the texture-based approach is similar to the Shear-Warp implementation but takes advantage of the capabilities of GPUs.

Implementation using 2D Textures

For GPUs that only support 2D textures the polygon layers have to be aligned with an orthogonal axis of the volume, this is called *object-aligned* slicing. Three sets of 2D texture slices are required, one for each orthogonal axis, this is to ensure that if the viewing direction changes there is always a set of data that is close to perpendicular (Rezk-Salama & Engel 2000, Engel & Ertl 2002).

There are several limitations of the standard 2D texture based method. The first is that due to three sets of the volume data being stored there is a lower limit on the size of the volume that can be rendered. Further, when mapping the textures to the polygonal layers only bilinear interpolation is used which can cause visual artefacts (Rezk-Salama & Engel 2000). By inserting intermediate slices between existing slices the visual artefacts can be minimised at the expense of increased memory usage, or processing time if computed on the fly. Finally visual glitches can occur when the view direction is changed, forcing a different set of axis-aligned slices to be rendered.

Methods have been developed to remove the need for storing multiple copies of the volume. In (Engel & Ertl 2002) a single stack of object-aligned slices is used and then *viewport-aligned* slices are created by interpolating along multiple object-aligned slices. Next, trilinear interpolation is used by performing linear interpolation across the slices on the GPU. This can reduce some of the visual artefacts of the standard algorithm at the expense of increasing the complexity.

Implementation using 3D Textures

Instead of storing the volume as three sets of 2D texture slices, with appropriate hardware, the volume can be stored as a single 3D texture (Cullip & Neumann 1993). Unlike the 2D texture approach the polygon slices that will be blended and composited are aligned parallel to the viewing direction. This eliminates visual issues when the viewing direction is changed and removes the need for multiple sets of textures to be stored.

A 3D texture allows the use of trilinear interpolation, which results to a higher degree of image quality than the standard 2D texture based method. Due to the entire texture being stored in one form and the use of trilinear interpolation, to increase the sample rate only the number of polygon layers need to be increased. As new texture slices do not need to be created this method does not require significantly more memory than the 2D texture approach.

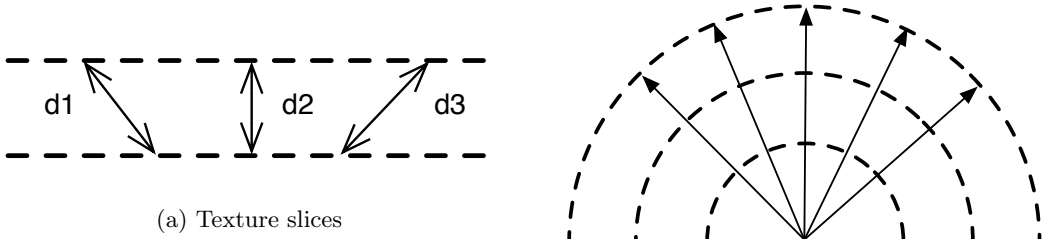
Sampling with Textures

In the 2D and 3D texture approaches, due to to the polygons being rendered parallel to the viewing direction when perspective projection is used, the distance between each slice along the viewing direction is not consistent (Kruger & Westermann 2003). Figure 2.10a shows the distance between each slice and how this differs depending on the viewing angle. The distance between each slice corresponds to the sample rate, therefore with texture mapping the sample rate is not consistent producing inaccurate images. The amount by which the distance changes depends on the field of view (FOV); with a small FOV the error decreases but is not removed. To eliminate the error, rather than use parallel polygons, a spherical shell rendering can be used, centered around the camera origin as shown in Figure 2.10b. This ensures that the sample rate is consistent, however this comes with an increased cost for rendering the more complex shell geometries (LaMar et al. 1999).

2.10.3 Order Independent Techniques

Whilst the methods discussed above try to evaluate the volume rendering equation, other methods of direct volume rendering algorithms use other ways of defining how voxels in the volumetric data are displayed. As might be suggested by their name, order independent techniques produce the same image regardless of front-to-back or back-to-front projection.

An alternative to DVR is Maximum Intensity Projection (MIP). Unlike full volume rendering, for each ray sent into the volumetric data this method returns only the maxi-



(a) Texture slices

(b) Shell slices

Figure 2.10: Polygon geometries using Texture Mapping.

mum density along the ray. As this only retrieves a single value there is no requirement for a potentially complex transfer function. However, as only a single density is returned for each ray a large portion of the volume data is discarded. This can be a concern if features in the volume with low densities need to be visualised as these will be ignored in favour of a higher density feature. Therefore MIP is often used to display skeletal structures with a high density value (Nelson et al. 2008).

X-Ray like renderings can also be supported by volume rendering. In this case the values along the ray are summed up and then averaged to produce an intensity value for the ray. Compared to full volume rendering this does ensure that each voxel in the data set has an equal probability of being displayed in the output image. However, as all voxels are potentially visible it may not be suitable if only certain parts of the volume need to be displayed.

An issue with both MIP and X-Ray renderings is that due to the ordering of the samples being the same regardless of viewing direction, occlusion is removed as a depth cue. In data sets that contain multiple overlapping features such as angiography imaging, this can be a particular problem as any spatial relationships between features will not be apparent. Modifications of the order-independent algorithms have been used to overcome some of the said issues, these include using chromostereopsis to differentiate different depths of features (Ropinski et al. 2006), as well as modifying the contrast of the scene to simulate aerial perspective (Kersten 2006). Stereoscopic displays have also been used to increase the perception of depth with order independent techniques (Poston et al. 1995, Mora & Evert 2004). Another method to increase the perception of depth is by rotating the volume, thereby introducing the kinetic depth effect (Boucheny et al. 2009, Kersten et al. 2006).

2.10.4 Domain Rendering

The methods discussed above deal with the presentation of the data in the spatial domain, either by iterating over pixels in the image or by iterating over voxels in the data set. Another way of rendering is to transform the data into a different domain, for example the Fourier or the wavelet transform or a compression algorithm.

The seminal studies on Fourier volume rendering Malzbender (1993), Dunne et al. (1990) independently designed a method whereby the volumetric data is converted into the frequency domain via the use of a Fourier transform. Using the projection-slice theorem (Levoy 1992), an inverse 2D Fourier transform is applied that results in a 2D projection of the data set in the frequency domain. Whilst the rendering time for the methods decreased from $O(N^3)$ for the simple *image order* and *object order* methods, to $O(N^2 logN)$, the methods only allowed for X-Ray like renderings and so occlusion and shading were not possible. Further, the initial methods were limited to only orthogonal renderings.

Shading and depth cueing were added to these methods to overcome the aforementioned issues (Levoy 1992, Totsuka & Levoy 1993). Relatively recent improvements to frequency domain rendering have permitted the use of full volume rendering thereby removing some of the earlier restrictions on visualisation type (Artner et al. 2005). A significant advantage of rendering in the frequency domain, other than decreasing rendering time, is the potential to use the sinc function to produce the final rendering.

2.10.5 Ray casting

Ray casting, as introduced by Levoy (1987), is an image-order method used to evaluate rays from the eye point through the volume, based on a discretisation of the volume rendering integral.

For each pixel in the display, either a single ray, or multiple rays if sub-sampling the scene, is traversed through the volume. Samples are taken along the ray at equal distances. At each sample point along the ray an interpolation filter is used to transform the discrete data set into a continuous form. Colour and opacity for the sample point can then be produced via the use of a transfer function and the gradient at the point may be calculated depending on the use, or not, of a shading function. The colour value is then *composited* with the existing colour and opacity for the ray. When the set number of samples have been taken, or the ray has passed fully through the volume, the accumulated colour and opacity are used as the pixel value (Levoy 1988).

The rendering time required to produce an image is directly related to the number of pixels in the image, the distance between samples and the choice of sub-sampling method. Classically, this has caused the ray casting method to be rather slow leading to the creation of some of the now competing volume rendering algorithms such as Shear-Warp and Splatting. In order to reduce the rendering time, a number of modifications can be made that include:

Early Ray Termination This is a simple modification to the raycasting algorithm. If a ray traverses through the volume and accumulates a high opacity, typically 0.95, the ray is terminated. Depending on the transfer function used and the data set this can decrease the rendering time (Levoy 1990).

Spatial Data structures Typically in a volume dataset a large proportion of the voxels will be set to 0. When casting rays through the empty voxels the algorithm is wasting processing time on areas that will not contribute to the final rendered image. By using a spatial data structure such as an octree that can encode regions, rays can skip over the blank voxels and so further reduce the rendering time required whilst not impacting image quality (Engel et al. 2006).

Adaptive Sampling In areas of the volume with small variation it is not necessary to have a high sample rate. By adaptively changing the sampling rate during traversal, samples can be taken more frequently where they are needed and less so when the data is consistent.

Template Based Rendering In orthographic viewing each ray will share the same direction and details regarding sample positions. By storing these in a *template* time can be saved having to recalculate the sample positions and directions that all of the rays will share (Yagel & Kaufman 1992, Yagel & Ciula 1994, Lee et al. 1997). However due to this method exploiting the coherence between parallel rays, it is not practical for perspective projection where rays will not be parallel and therefore not share sample positions or directions.

GPU Implementation Wit the introduction of general-purpose computing on GPUs raycasting can be implemented directly on graphics hardware. Many of the above speed improvements can also be implemented on the GPU, further decreasing rendering time.

2.11 Conclusions

In this chapter the background on volume rendering has been reviewed. Volume rendering itself is a broad topic with a range of implementations, volume formats, parameters and choices to be made regarding how the volumetric data is presented to the user. It is evident that choices in the volume rendering pipeline including sampling strategy, reconstruction filter and classification can have a significant impact on the final rendered image. Each parameter in the algorithm must be considered carefully and therefore there is the need to understand how the choices affect the understanding of the images produced.

Chapter 3

Human Vision and Depth Perception

This chapter provides an overview of the human visual system (HVS) including relevant sensitivities, as well as providing an overview of the literature in regards to depth perception and their realisation in 3D displays.

3.1 Depth Perception

The HVS is capable of using a wide range of cues to perceive depth. Figure 3.1 shows a taxonomy of the depth cues. Depth cues that can be presented on a 2D display or image are known as pictorial or psychological depth cues. Binocular cues arise from angular distance between the two eyes and oculomotor, or physiological cues, use feed back from the muscles in the eyes (Blake & Sekuler 2006).

3.1.1 Pictorial Depth Cues

Pictorial or monocular depth cues can be presented in a single 2D image and can be perceived with only a single eye. Due to this they have long been used in paintings and computer graphics to produce a sense of depth into an otherwise flat image.

Linear Perspective

In the real-world, objects at a distance from the viewer will eventually converge to a point. This is most obvious with parallel lines such as train tracks, as the tracks recede in the distance the tracks will converge to a point of the image. The size of objects also change

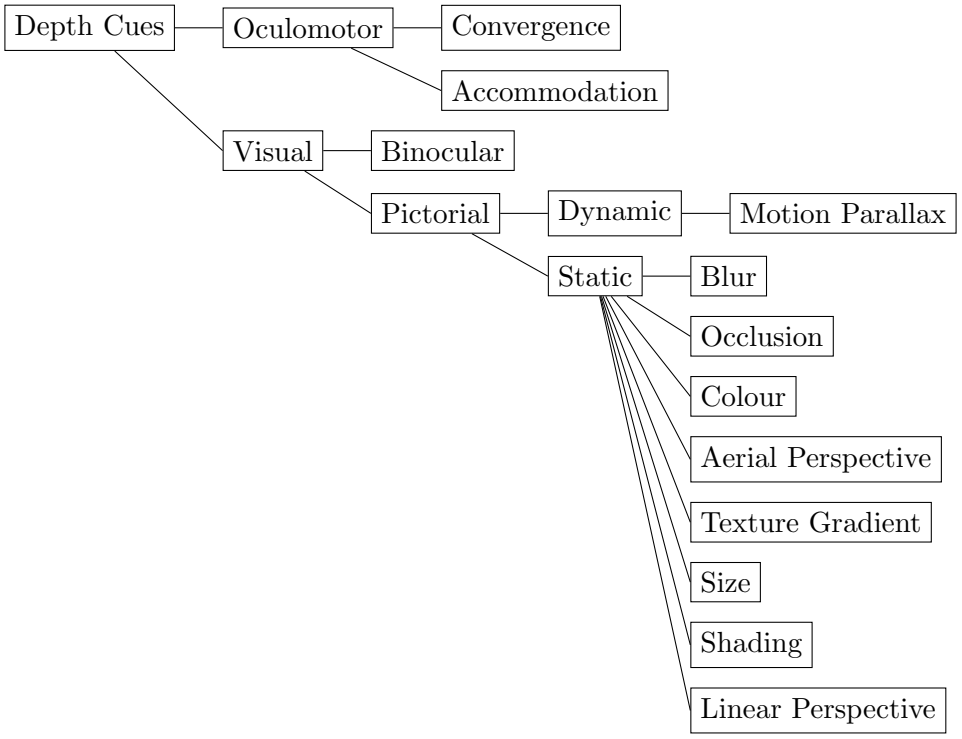


Figure 3.1: Taxonomy of depth cues

as their depth increases with objects appearing smaller (Goldstein 2002). As Figure 3.2 shows this cue was particularly troublesome for pre-renaissance era artists who had yet to fully-determine how perspective could be duplicated in drawings and paintings. Alberti, a renaissance era artist, is credited with devising one of the first methods for faithfully reproducing perspective projection in paintings by tracing the image of a scene as it would be presented in a window (Palmer 1999).

Shading and Shadows

Here we make a distinction between shading and shadows as a depth cue. Shading darkens areas on a surface according to the amount and direction of the light reflected from the surface (Palmer 1999). Formally it is defined as the variation in radiance caused by surface orientation or specular differences. Shading can greatly change or alter how we perceive the structures of shapes. By shading spheres differently, either convex or concave shapes can be perceived and it is believed that the reason is due to an unconscious assumption that the light source causing the shading is located above the viewer (Ramachandran 1988).

Shadows are caused by objects occluding the path to a light source. Depending on the

Figure 3.2: Pre-renaissance era painting depicting the difficulty in reproducing accurate linear perspective (*Entry into Jerusalem*, Pietro Lorenzetti 1320)

position of a shadow a large amount of spatial information can be provided. As can be seen from Figure 3.3 shadows can provide a cue as to the relative depth of objects in a scene.

Size

When the relative size of an object is known an estimate can be made regarding the depth of the object. For example, given two objects that have the same size, if one of these objects is presented as being smaller, then it is implied that it is further away than the larger object.

Similarly, we can use familiar sizes to deduce some level of relative depth ordering. If an image shows an elephant and a mouse at the same size, using our familiarity of the sizes of the two animals we can deduce that the elephant must be behind the mouse. If the elephant was closer than the mouse then it would be presented considerably larger than the mouse (Goldstein 2002).

Texture Gradient

Textures or patterns on a surface can also be used as a perception of depth. The denser the patterns are the further away they appear to be from the observer. The effect was

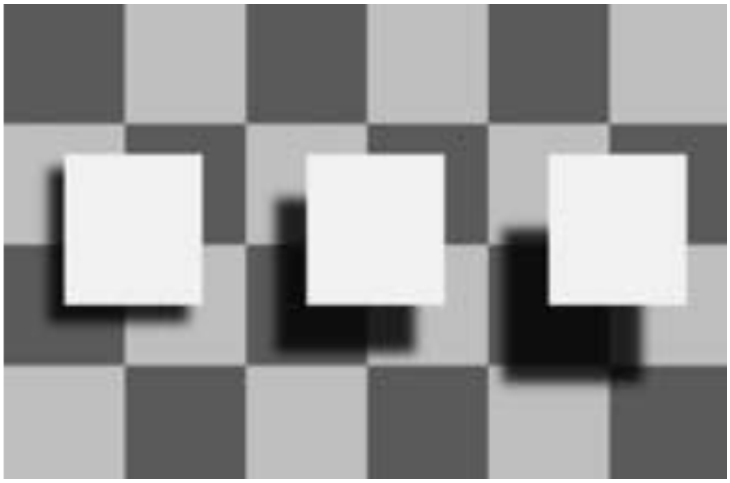


Figure 3.3: Different positions of the shadows provide different perceptions of depth for each of the squares (Palmer 1999)

described by Gibson (1950) in his work encompassing the visual perception. Textures can also be used to provide cues as to the orientation and curvature of a surface (Palmer 1999).

Aerial Perspective

At a far enough distance light gets scattered by particles or moisture in the atmosphere. This adds a level of fuzzyness or blur to objects that are far in the distance.

As well as a decrease in sharpness, objects in the distance will appear with an increasing blue tint. This is due to the longer wavelengths of light being scattered in the atmosphere, short wavelengths that correspond to the colour blue are not scattered to such a degree. This is the only source of depth in which the effectiveness increases with the distance from the user due to the increase in density of particles (Cutting & Vishton 1995).

Colour

Colour can produce a perception of depth known as chromostereopsis. Each colour has a different size wave-length, the focusing of the different wavelengths within the eye can cause light such as red with a long wave-length and blue with a short wave-length to appear at different depths. However, the effect of chromostereopsis is not guaranteed and is dependant upon the user. With red and blue text on a black background up to 60% of people will see the red text as being in front of the blue text. The other 40% will either see blue in front of the black or no depth effect (Ware 2004). This makes the effect of chromostereopsis unpredictable.

Occlusion

When one object A obscures all or part of another object B, then we know that object B is behind object A. Whilst providing relative depth judgements, this cue can not give absolute depth judgements. However, unlike some other depth cues, the effectiveness of occlusion is constant regardless of the distance from the user (Cutting & Vishton 1995).

Blur

When our eyes accommodate on an object at a distance, only objects within a small region around the object appear sharp. Objects in front of this distance and behind it are displayed with some levels of blur; the precise equations governing the amount of blur received are dictated by the thin lens theory (Mulder & van Liere 2000). Due to blur being added both in front of and behind the region of focus, the sign, or ordering, of the relative depths is not available. This gives the initial impression that blur is not a useful cue due to lack of depth ordering. However, the differences in blur between regions and borders of the regions can be an effective cue and can provide signed information (Pentland 1987, Mather & Smith 2002).

Motion Parallax

Motion parallax is another strong depth cue and can help in determining 3D spatial relationships (Holliman 2005). Unlike the static pictorial depth cues this relies on either the observer moving their head, or the observed item having some horizontal motion. These are sometimes referred to as *head-motion parallax* and *object-motion parallax* respectively (Glassner 1995). If we assume that the observer is moving from left to right then objects close to the observer will move at a greater distance across the retina and so be perceived as moving faster. Objects at some further distance will be projected as moving slower across the retina. The different speeds of the near and far objects can then be used as way of determining the depths of objects (Goldstein 2002, Reichelt et al. 2010). Initial experiments performed by Gibson et al. (1959) investigated whether the movement of light when entering the eye has an affect on relative depth ordering and later tests confirmed that motion parallax by itself can provide the impression of depth without any other depth cues available (Rogers & Graham 1979). The effectiveness of this cue can exceed the 10 meters of binocular disparity (Nagata 1984), as detailed further in Section 3.1.4.

3.1.2 Oculomotor Depth Cues

When we perceive an object in the real-world our eyes naturally accommodate or focus on the object and converge or angle themselves towards the object that is to be in focus (Reichelt et al. 2010, Mcallister 1993). The oculomotor cues derive from the feedback from the muscles that perform the accommodation or convergence (Blake & Sekuler 2006).

The usefulness of these cues for the perception of depth is limited, as concluded by a review of the results of accommodation and convergence for depth perception (Blake & Sekuler 2006). Whilst it was found that convergence can act as a reliable depth cue for distances of less than 20 feet, after this distance the eyes are parallel and so the angle of convergence approaches 0. The review further found that accommodation was largely inaccurate in determining depths. A study by Richards & Miller (1969) reports that up to a third of people may be unable to use convergence as a depth cue, making it particularly unreliable.

When using stereoscopic displays the natural relationship between accommodation and convergence is broken. The eyes will converge to the point that is displayed in depth however they will accommodate on the plane of the screen. This causes a conflict and is a cause of visual fatigue and discomfort when viewing stereoscopic images (Ukai & Howarth 2008, Lambooij 2007).

3.1.3 Binocular Depth Cues

Binocular depth cues cannot be presented in a single 2D image, instead a stereoscopic display is required that is capable of displaying two or more images to the viewer at the same time. The HVS contains two forward-facing eyes that due to the *Interocular Distance* (IOD) of approximately 65mm for an adult see different overlapping perspective views of the world that create a *binocular visual field* (Ware 2004, Blake & Sekuler 2006). The depth cue that arises from the binocular view of the world is called *stereopsis*, in Greek meaning 'solid seeing' (Blake & Sekuler 2006).

Retinal disparity is the lateral separation between corresponding points in the views projected onto the two retina. When the eyes fixate on an object in the scene, the points corresponding to that object in the retina will have zero retinal disparity and are said to be *corresponding retinal points* (Goldstein 2002). Objects that are nearer and further from the fixated object will have some retinal disparity, with those nearer to the viewer having *crossed disparity* and those further away from the viewer having *uncrossed disparity*. If

the retinal disparity between the two images is too large then *diplopia* will occur, that is, the two images will be unable to be fused and the object will be seen as two separate images (Ware 2004).

A small percentage of users are unable to perceive binocular disparity and so suffer from what is called *stereoblindness*, or perceive the inverse affects of depth when presented with images that use binocular disparity alone (Richards 1970). Results from a trial of 150 people show that 4% of the sample size were unable to use binocular disparity and a further 10% reported incorrect depths during the study (Richards 1970). Therefore pictorial depth cues are the main methods that these groups of people can use to perceive depth.

Stereo-acuity

Stereo-acuity is defined as the smallest angular difference in depth between two objects that can be determined by a viewer and is represented by the symbol δ. It is given in angles as this allows the value to be constant regardless of distance to or between the actual points (Holliman 2005). Reported values are given as arc minutes ($'$) equal to $\frac{1}{60}$ of a degree and arc seconds ($''$) $\frac{1}{60}$ of an arc minute. In Figure 3.4 consider the points n and f that are perceived as being at disparate depth planes. The disparity in degrees between them is $\delta = (\beta - \alpha)$ where:

$$\beta = 2 \times tan^{-1}\left(\frac{IOD}{2D_f}\right) \tag{3.1.1}$$

and

$$\alpha = 2 \times tan^{-1}\left(\frac{IOD}{2D_n}\right) \tag{3.1.2}$$

where IOD is the interocular difference, D_f is the distance to point f and D_n is the distance to point n.

There are a number of different stereo-acuity tests that can be used to generate values for the smallest depth that can be perceived. Broadly speaking such tests can be segregated into *real depth tests*, those that generate binocular disparity due to physical differences of depth in the real-world and those that generate differences of depth due to the use of horizontal parallax of two views that are either printed or displayed on a 3D display. A typical example of a real depth test includes the Howard-Dolman style test (Howard 1919). Using two cylinders separated horizontally at a distance from the participant, in each test one of the cylinders is placed in front of the other cylinder. The participant is required to state which of the cylinders is in front of the other and if the participant judges correctly

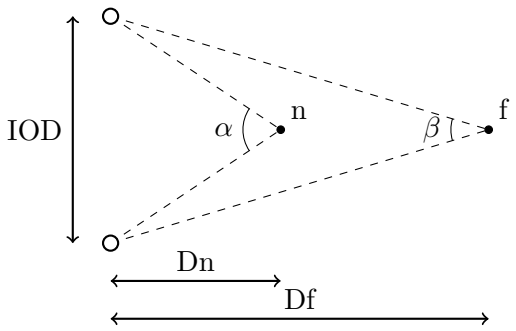


Figure 3.4: Binocular geometry.

then the distance between each of the cylinders will decrease. The smallest depth the user can perceive is the smallest depth at which they are reliably able to state which cylinder was in front of the other.

An early experiment using the Howard-Dolman test determined that pilots were capable of detecting as little as $1.8''$ of disparity. Other studies have reported similar results using random dot stereograms (RDS) and stereoscopic displays (Julesz et al. 2006, Yeh & Silverstein 1990). Whilst these are the smallest values of stereo-acuity that can be detected by a user, Coutant & Westheimer (1993) reports that approximately 97% of 183 users, could perceive a disparity of $2.3'$ and further that 80% of these users could see a disparity of $30''$.

Fusion Limits

The upper limits of fusion with stereoscopic displays before diplopia occurs have been explored with the results from Yeh & Silverstein (1990) showing an upper limit of $27'$ for crossed disparity and $24'$ for uncrossed disparity for stimulus that is displayed for a brief duration of 2 seconds. Yet with longer durations of stimulus the limits can increase to 4.93 degrees and 1.57 degrees for crossed and uncrossed, respectively. Pastoor (1995) found that disparities of only $35'$ of arc could be displayed before visual discomfort became an issue. Others have suggested that to alleviate possible discomfort caused by excessive disparity, a limit of 1 degree should be adhered to (Lambooij 2007). In determining a comfortable range of depth limits Jones et al. (2001) suggests perceived depth that is equivalent to $\pm 60mm$, even though it is likely that many people can still perceive depths at greater than this limit. In any case it is clear that there is a limit to the amount of disparity that can be displayed before the images can no longer be fused and that even if fusion occurs uncomfortable side effects can still be produced.

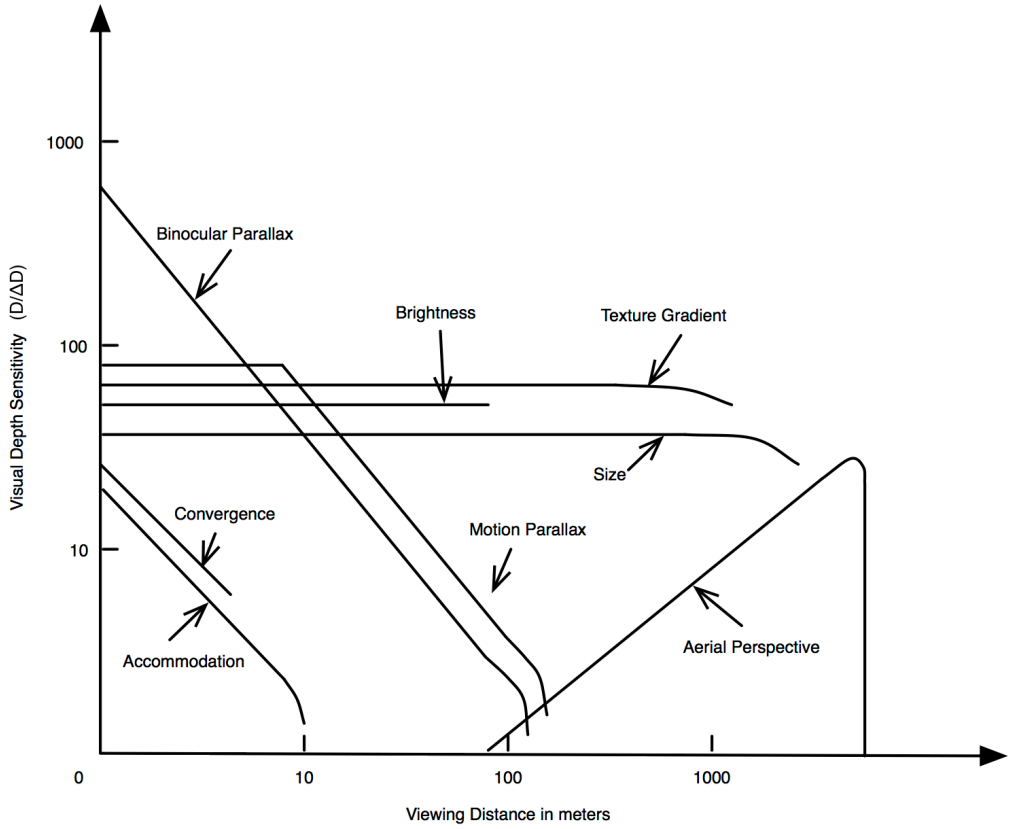


Figure 3.5: Effectiveness of depth cues at different distances against depth sensitivity, where D is viewing distance and ΔD is depth threshold (Nagata 1984)

3.1.4 Effectiveness of Depth Cues

Not all depth cues are effective at the same distances or assist in all depth tasks. Cutting & Vishton (1995) and later Nagata (1984) performed depth experiments to judge the effectiveness of convergence, accommodation, binocular parallax, motion parallax, brightness, gradient texture, size and aerial perspective. The results in Figure 3.5 show the similar values between motion parallax and binocular parallax as well as the limited ability to facilitate convergence and accommodation which are only effective up to 10 metres. Note that aerial perspective is the only depth cue that increases in effectiveness with distance. Cutting & Vishton (1995) split the depth cues into three areas of effectiveness; *personal space* that is within 2 meters of the user, *action space* that is within 2 to 30 meters of the user and *vista space*, 30 meters and beyond from the user.

Combination of Depth Cues

From the prior sections it is clear that the human visual system can use a wide variety of depth cues to determine ordinal depth, absolute depth and size. Away from experimental conditions, the human visual system uses combinations of depth cues to provide an accurate visualisation of the scene. Therefore a problem that has been researched in the literature is how the human visual system combines these cues.

Three theories are dominant in how the human visual system combines multiple depth cues. The *Weak fusion* model suggests that each of the cues present a *depth map* (Palmer 1999) in isolation, these are then combined via some weighted linear average to produce the final depth map of the scene. Several studies endorse this method (Cutting & Vishton 1995, Johnston et al. 1994, Clark & Yuille 1990) where the modularity of the depth cues allows each one to be tested in isolation. However such a theory has been criticised (Landy & Johnston 1995), in that the weak fusion model suggests all depth cues provides the same type of information which is then combined. However, this is not the case as for example motion parallax is able to provide absolute depths where as occlusion provides only relative depth judgements, which would require the values to be somehow converted to the same units and then combined. Further, from studies conducted by Nagata (1984), Landy & Johnston (1995) it has been demonstrated that depth cues have different effective ranges and this would require the weighted averages to be dynamic within the scene.

The alternative model, *strong fusion*, puts forth the theory that the depth cues are not treated in isolation and may interact with each other freely so as to produce a final depth for the scene(Johnston et al. 1994).

Landy et al suggest a half way theory known as *modified weak observer* or *hybrid fusion*. The depth cues are still treated in isolation to present some depth estimate, however they can also interact in a limited fashion. An example is that a depth cue that can only provide relative depth judgements is promoted by another depth cue that can introduce the missing absolute depth values. This then allows both cues to produce absolute depth values.

Conflicting Depth Cues

These models suggest ways depth cues are combined, but not what happens if depth cues conflict. In natural scenes this is rarely an issue, however in stereoscopic displays conflicting depth cues can be present.

One such example is known as *frame cancellation* (Ware 2004), where occlusion dictates that the surround of the screen is in front of the stereoscopically presented content yet binocular disparity suggests that the stimulus is in front of the screen. This can cause an unsettling visualisation (Lipton 1997).

In the case that cues do conflict, there are a number of ways they may be handled by the HVS, either one of the cues will be dominant and suppress the weaker cue, both cues will be combined to some degree or the perception of the scene will fluctuate due to neither cue having a dominant strength (Drascic & Milgram 1996).

3.2 3D Displays

There are a range of stereoscopic systems now available and they can be categorised under different headings. We will follow the literature that splits the displays by the number and type of views that a display is able to present to a user. These are *binocular* systems that present two views to the user, *multiview displays* that provide multiple views of a scene allowing viewing from different directions and finally *full-parallax* displays that allow a scene to be seen from any viewing angle. We will briefly review the types of displays, with more information regarding exact hardware specifications available in Holliman (2011), Reichelt et al. (2010) and Pastoor & Wöpking (1997).

3.2.1 Binocular Displays

Binocular planar displays present only a single pair of left and right images to the user. The stereoscope of Wheatstone (1838) falls under this category. There are two methods for displaying the views to each eye. *Time parallel* displays present each view to the left and right eye at the same time. In contrast, *time sequential displays* present each view one after the other (Holliman 2011). Time sequential displays do require that the views are switched at greater than 58 Hz to ensure that the highest intensity light is represented without flickering (Hecht & Smith 1936).

Stereoscopic Displays

Stereoscopic displays require a form of head wear to be worn by a user so that the light for each view can be selectively filtered to the correct eye. These can be passive, as in the case of anaglyphs and polarized glasses, or active in the case of head mounted displays and

shutter glasses (Reichelt et al. 2010). In the past the most frequent method for displaying stereoscopic images both in printed form and displayed on screen was via a method known as *anaglyph*. This works by filtering two views of a scene to each eye by wearing a special pair of passive glasses. Commonly red/cyan glasses are used with each colour lens filtering only the respective colour channels to each eye. As well as being used in the first generation of 3D cinemas it is still in common usage today as the images can be printed or displayed on any standard colour monitor and the glasses are easily available. The downside is that as the lenses use colours to filter the scenes, full colour images cannot be used and the coloured filters are unable to completely block the image meant for the other eye, resulting in an issue called crosstalk, defined in Section 3.2.4.

Where as anaglyph images do not require any particularly special display hardware, the use of *polarised glasses* do but are able to present full-colour images. A filter is applied to the screen that is used to change the angle of the polarised light in the case that a linear polarised filter is used, or the direction of the light in the case that a circular polarised filter is used. Each lens in the glasses will then only permit either one of the angles or directions of light to pass allowing each view to be directed to the correct eye. A potential issue with this display method is that due to the use of polarised filters light intensity can be lowered and this should be taken into account when using this hardware (Holliman 2011, Sexton & Surman 1999). Also, in the case that only a single display is used for both views then the resolution of the views will necessarily be half of the full resolution of the display.

In contrast to anaglyph and polarised glasses, the use of *shutter glasses* requires active glasses that alternatively block the view of the display from reaching one of the eyes. As the display alternatively blocks one of the views and does not show them together, this is a *time sequential* method and requires that each pair of glasses is synced with the display to ensure that each lens in the glasses admits the correct view at the correct time. In the past mechanical shutting devices were used to block the light however these were found to be too slow and bulky (Lipscomb 1989). Now liquid-crystal systems are used as these provide the necessary speed to ensure that the appropriate frame rate is matched for each eye. Unlike stereoscopic displays that use a single screen and time parallel methods that show both views together, here the full resolution of the display is available for each view.

A particular type of display that can be used to present stereo is the head-mounted display (HMD). This is a wearable device that presents two views directly to each users

eye commonly by the use of small LCD displays placed directly in front of each eye. Head tracking is commonly used with this type of display allowing a user to move their head as desired. A user is free to move their head as desired without impacting the view of the image that is presented and so head tracking is a common addition (Pastoor & Wöpking 1997).

Autostereoscopic

Autostereoscopic displays present a 3D image without requiring users to wear possibly cumbersome head wear. There are two flavours of two-view autostereoscoic display. Those that use two displays to present the user with the 3D image and those that use a single divided display. Both types are *time parallel* methods as a user must be able to see both views at the same time.

Single screen autostereoscopic displays require a method of dividing the display into two separate views that are presented individually to the user. Parallax barriers can be used to block light from being seen at different angles. Other methods include lenticular lenses that angle the light towards each eye and micropolarisers that can act as switchable parallax barriers (Mcallister 1993, Holliman 2005). Each method divides the resolution of the display into two, with alternating columns of pixels belonging to different views. Due to the alternating columns a user must be sitting in precisely the correct location to view the stereoscopic image, if the user is not then they will receive an inverse stereoscopic image, or *pseudoscopic* image. There is only a 50% chance that the user will be sitting in the correct position if no markers or assistance is provided (Dodgson 1997, Holliman 2005).

Twin-screen autostereoscopic displays use two displays, one for each eye, which must then be directed towards each eye. The main benefit of these displays is that a full resolution image can be presented to each eye.

3.2.2 Multi-view Displays

These are autostereoscopic displays that present more than one view of the scene and can allow for multiple users to 'look-around' the stereo scene. There are three different methods for creating a multi-view display that depend on how the views are created (Dodgson 1997). Spatial multiplex displays split a single display into multiple views, using the same methods as for stereoscopic displays, either via lenticular, or parallax barrier technologies

(Holliman 2011). Another method uses multiple projectors to create each view and whilst it is able to produce scenes with the same resolution of the projectors used there can be a considerable cost involved. Finally a time-sequential display rapidly alternates between the different views to be displayed however this requires a screen with a rapid frame rate.

3.2.3 Full-Parallax Displays

A class of displays known as *volumetric* displays provide both vertical and horizontal disparity. These are fundamentally different from other planar autostereoscopic displays in that depth is not the result of the disparity between points on a display. Each display element, also called a voxel in this context, emits light from the true x,y,z position in space (Favalora 2005). As the points are presented at their true position in space, accommodation and convergence cues are kept consistent and so the issues of unrealistic accommodation described by Lambooij (2007) for planar stereoscopic stereoscopic displays are not a problem in this case. There are several different types of volumetric display including swept volume. Where light is directed onto a rapidly rotating surface and as long as the refresh rate is high then the 3D scene can be perceived. Another choice is a static volume display whereby the major part of the display is kept static and the voxel elements within this zone emit light (Favalora 2005).

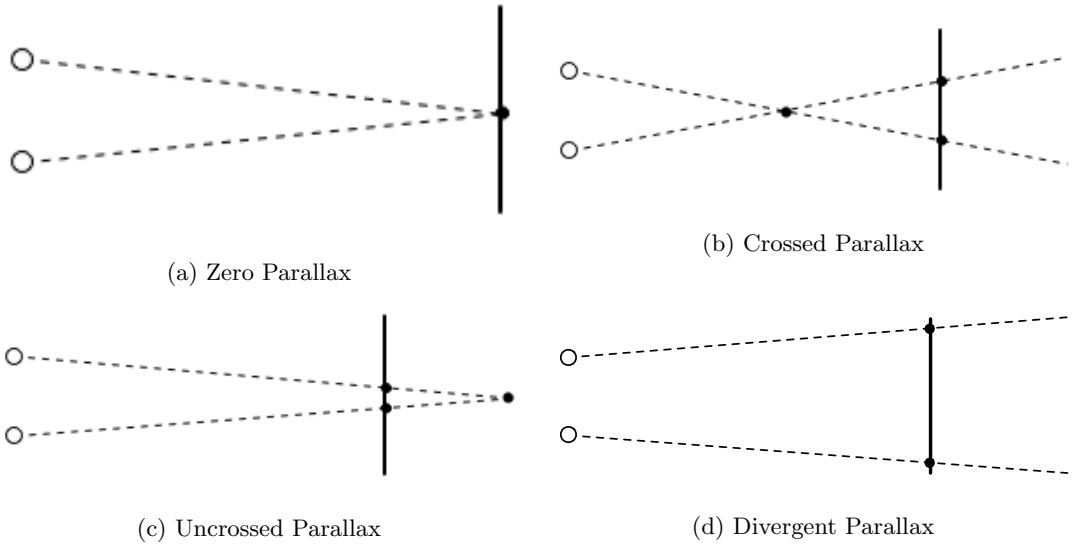
3.2.4 Considerations

When using (auto)stereoscopic displays, there are a number of considerations that should be taken into account. Some of these are parameter based, such as the camera geometry, whilst others are due to properties of the screen and limitations of planar displays, such as accommodation.

Parallax

When viewing with stereoscopic displays the horizontal difference between matching points in a scene is called the *parallax*. The types of parallax that can be induced with a stereoscopic display are *zero parallax*, *crossed parallax*, *uncrossed parallax* and finally *divergent parallax*.

Figure 3.6 shows the converging angle of the different parallax types. Objects with zero parallax will lie on the screen plane and will have no depth, objects that have crossed parallax will be perceived as being in front of the screen plane, where as uncrossed will



(a) Zero Parallax

(b) Crossed Parallax

(c) Uncrossed Parallax

(d) Divergent Parallax

Figure 3.6: Different parallax types for a stereoscopic display.

be perceived as being 'inside' the screen. Note that the final type of parallax, divergent, is rarely used in practice. This is because it forces the eyes to separate at some distance greater than the Interocular Distance (IOD). This type of parallax is not found in the real-world and so should not be used when rendering stereoscopic scenes (Lipton 1997).

Resolution

With 3D displays the number of depth layers that can be presented is a function of the horizontal pixel count of the display. In Figure 3.7 we can see the depth layers are produced by the polyhedrons formed between the IOD and horizontal resolution of the display. The polyhedrons are called *stereoscopic voxels* and any objects that are contained in the same stereoscopic voxel will be perceived as being at the same depth. The HVS itself is capable of perceiving a considerably higher number of depth layers than most stereoscopic displays are able to produce (Holliman 2005).

Camera Setup

When generating stereoscopic images there are two standard camera methods used, known as toed-in and parallel asymmetric. Figure 3.8 shows diagrams of the camera geometries.

The toed-in method uses two cameras that are separated by a distance but are pointed at a common focal point. Although simple, this method can cause two type of distortions: keystone distortions and depth plane distortions (Docherty & Koch 1993). Keystoning

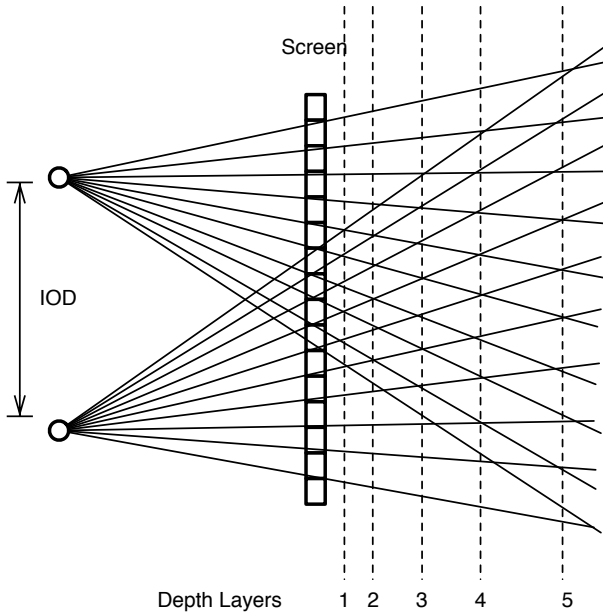
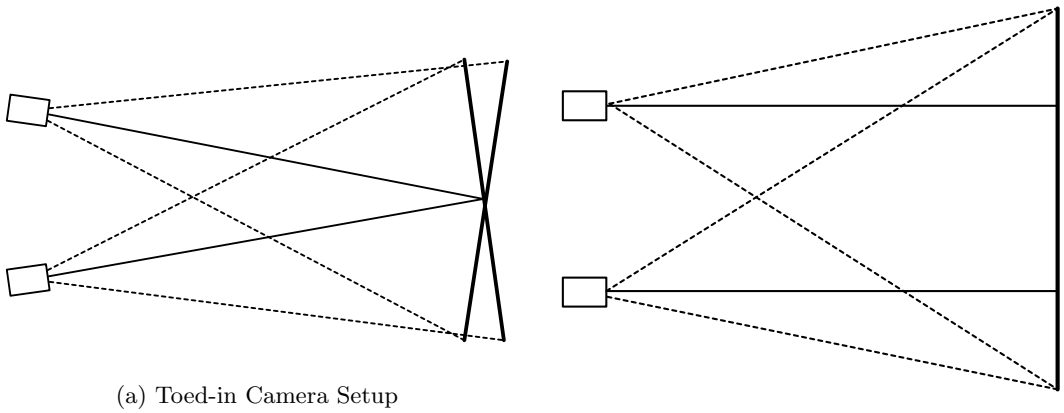


Figure 3.7: 3D area defined by Stereoscopic voxels adapted from (Hodges & Davis 1993).



(a) Toed-in Camera Setup

(b) Parallel Asymmetric Camera Setup

Figure 3.8: Camera setups for stereoscopic geometry (Docherty & Koch 1993).

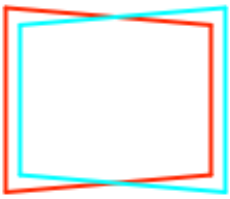


Figure 3.9: Keystoning artefacts (Docherty & Koch 1993).

occurs due to perspective projection where the image planes form a trapezoidal shape shown in Figure 3.9. This leads to vertical and horizontal parallax, with the greatest parallax induced in the corners of the image. Studies into vertical parallax have shown that 4'-7' represents the limit of vertical disparity for the HVS (Nielsen & Poggio 1984), beyond this value diplopia can occur. In most cases it is prudent to avoid the risk of vertical disparity being introduced without need. This effect can also cause depth plane curvature where objects in the centre of the display appear closer than objects at the corners of the display.

The alternative is to use parallel cameras with asymmetric frustums that have view directions that are parallel to each other. This does not create the vertical disparity issues of the toed-in method.

Accommodation and Convergence

In the real-world the accommodation and convergence of the eyes are linked. If the eyes are accommodated on an object then they will also be converged on the same object. For planar stereoscopic displays the eye is still converged on the stimulus presented but the planar display requires us to accommodate on the screen (Lambooij 2007). This forces us to override the natural actions and can cause discomfort (Reichelt et al. 2010). Stereoscopic and auto-stereoscopic displays will all suffer from this and although the affect can be weakened by limiting the depth used, it cannot be solved. With increased usage and user training with stereoscopic displays the link between accommodation and convergence can be weakened reducing the risk of discomfort and allowing greater depths to be presented (Docherty & Koch 1993).

Crosstalk

Crosstalk is regarded as one of the main issues that can impact stereoscopic image quality. Within the literature there seems to be some confusion between the terms of crosstalk and ghosting with the terms being used somewhat interchangeably. We define crosstalk to be the incomplete separation of the left and right views that causes light from one view to leak into the other. Ghosting on the other hand is the perceived doubling of the images that is caused by the problem of crosstalk. (Lipton 1997).

Crosstalk depends upon the hardware used to display the stereo images and can be directly calculated, where as ghosting is a subjective phenomena. In a perfect system there would be no crosstalk and so a user would not receive any percept of ghosting. For example, if a head mounted display or a stereoscope is used like the Wheatstone design then this is possible. In most standard stereoscopic displays it is however not possible to be completely crosstalk free (Lipton 1997). From recent tests in the medical domain it has been found that crosstalk levels of 2-6% can have a significant detrimental impact on the perceived quality of stereoscopic images (Albani et al. 2011).

In classic CRT stereoscopic displays, a large cause of crosstalk is due to the non-immediate decaying rate of the phosphors used to represent each pixel in the display (Yeh & Silverstein 1990). For polarization techniques, crosstalk has been reported at values as small as 0.1% (Pastoor 1995).

In most cases displays that have a low-level of crosstalk should be preferred. However, in the case of multi-view displays, Pastoor (1995) and Lambooij (2007) suggests that to alleviate the problem of view flipping, whereby there are jumps between images caused by different perspective views, crosstalk could be used to add a level of blur between them, creating a more seamless transition.

3.3 Factors that can Impact Depth Perception

Besides the technical factors related to the properties of the displays, factors such as aliasing, blur and the contrast of an image can also impede the perception of depth. The studies of how these can impact the perception of depth are relevant for this thesis as aliasing and blurring can be caused within stages of the volume rendering pipeline. Blurring can also bring about a decrease in contrast of the image.

3.3.1 Aliasing and Depth Perception

3D Displays are susceptible to the standard 2D aliasing artefacts that include *spatial* and *temporal* aliasing caused by limited resolution of the frame buffer of the graphics device. Further aliasing can occur due to the: limited number of depth planes, which are a function of the horizontal resolution of the display (Pfautz 2000, Hodges & Davis 1993, Jää-Aro & Kjelldahl 1997), sampling issues between the number of views in the stereoscopic display causing *intra-perspective* aliasing (Moller & Travis 2005) as well as issues caused by the combination of perspective projection and 3D displays (Pfautz 2000).

2D Aliasing and Depth Perception

In describing the impact of depth perception on 3D displays an understanding of 2D aliasing artifacts is required. This is because, when combined with 3D displays, 2D aliasing can be the cause of non-correspondence aliasing, whereby points cannot be matched correctly between the two views of a stereoscopic image, thereby degrading depth perception (Reinhart 1992).

A 2D image can be represented as a 2D continuous signal. The frame buffer of the graphics card forces the use of a discrete 2D signal which is then reconstructed into a continuous signal by the display device. 2D signals, just as with 1D and 3D in terms of volume rendering, must be sampled at twice the highest frequency contained in the source to be accurately represented. In static 2D images the limiting factor for the accurate resampling of the 2D signal is the resolution of the frame buffer in which the visual information of the image is stored prior to sending it to the display. If the image contains sharp boundaries between features then it is possible that an arbitrarily large number of frequencies is required, well above what any practical resolution can support. Therefore the sampling theorem cannot be satisfied and aliasing artefacts will occur. Such aliasing, called spatial aliasing, manifests in the image in various ways depending upon the exact nature of the feature. For silhouettes of objects this results in jagged edges that appear as step-like formations; long thin objects may appear as discontinuous segments and detailed textures can lose some of the finer detail. Worse still is the impact on small objects, as small features between two samples can disappear entirely (Crow 1977). In animated scenes spatial and temporal aliasing can occur and can result in jarring motion, reversal of direction and flashing of objects that can appear and disappear. Crow (1977) advises three methods to reduce the spatial aliasing artefacts; increasing the resolution; blur the

image or use area sampling. The first, increase the resolution, is perhaps the most obvious, however it comes with a limitation in that the resolution can only be increased up to a finite limit. Blurring the image will obviously reduce the sharpness of the features in the image. Finally, area sampling will produce a sample point value by averaging the values contained within some area. The latter is the preferred option, however it can still cause smoothing.

Extending this understanding, Pfautz (2000) performs experiments into how sampling can impact depth in 2D perspective images. In any 2D image, due to rounding, the maximum perceived error of a location at an (x, y) point is up to half a pixel in the horizontal or vertical dimension. When perspective projection is used, this rounding causes represented depths to differ from the actual depths. At distances closer to the user such an error in location is not significant, however at distances further away from a user an error of up to half a pixel can have a greater impact on the depth presented. Further, this error can lead to inconsistent sizes and proportions.

The perceptual effects of 2D image aliasing on depth perception are less clear. Booth et al. (1987) studied the performance of participants in a mental rotation task. A tetris-like block was rendered at different pixel sizes in aliased and anti-aliased scenes. Participants were required to subjectively state the quality of the image and then in the task based study they were required to count the number of individual cells that made up each of the blocks. The results of the task suggested that at low and high resolutions there was no considerable difference between aliased and anti-aliased images. However, the subjective scores showed that participants preferred the images displayed at the higher resolutions.

Stereoscopic Aliasing with Depth Perception

In stereoscopic scenes, aliasing can be broadly classified into two categories depending upon how it is introduced. *Inter-perspective* aliasing is formed by the finite number of views that can be displayed on a planar stereoscopic display. Specifically this form of aliasing is caused by insufficient views of a 3D scene that produces discontinuities during the viewing on the 3D display. This is a problem for auto-stereoscopic displays whereby users are free to move between the different views (Moller & Travis 2005). If the auto-stereoscopic display is treated as a signal, then the required number of views should be twice that of the highest frequency of the scene. In practice, the limited number of views causes crosstalk whereby an eye will receive two views of the scene at the same time

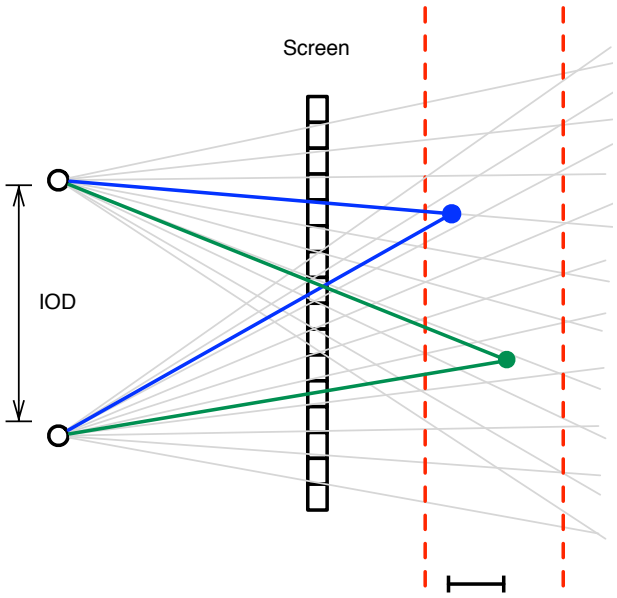


Figure 3.10: Two points at different scene depths can be given the same screen disparity and will have the same perceived depth.

thereby reducing the effectiveness and comfort of the viewed stereoscopic image.

The other category of aliasing, *intra-perspective*, is inherited from using 2D displays of a finite resolution. Due to the limited horizontal resolution there is a limit to the number of steps in depth that can be displayed, as discussed in Section 3.2.4. Whilst the finite number of depths is a key concern and cause of depth aliasing, a secondary problem also arises. As each stereoscopic voxel has a certain size, any points that lie within the voxel will necessarily have the same screen disparity and so will be displayed at the same depth. This means that given points A and B with different scene depths, they will have the same perceived depth when displayed. The problem is demonstrated in Figure 3.10.

In an attempt to understand depth aliasing, Jää-Aro & Kjelldahl (1997) explored the objective quality of depth in aliased and anti-aliased images in a high resolution workstation and a lower resolution HMD. Requiring users to assess the depth of a shape given the depths of two other shapes in the scene, it was found that at low resolutions pixels within the images did not align and that was a cause of misleading shape perception and ghosting, which had a negative impact on the ability to determine the correct ordering of the shapes. It was concluded that the stereoscopic display of the data was less crucial than the anti-aliasing of the images. In the tests using the HMD, there was little difference between the errors caused by aliased and anti-aliased images. It was thought that the low

contrasts produced by the HMD helped to naturally anti-alias the scene. This result aligns with that of a contrast based study that shows that the ability to perceive larger depth ranges can be helped by using low levels of contrast (Richards & Foley 1974).

In most cases it is advised that stereoscopic images are produced with perspective projection to avoid unwanted conflicting size perception due to *perceptual zooming*, where an object that is further away is perceived as being larger than an object that is closer to the user. However, when perspective is used with stereoscopic displays, sampling and aliasing effects can still be an issue. Pfautz (2000) determined that using perspective with stereoscopic displays can both complement the perceived depth and in some cases have a detrimental impact. This is due to the problem that the levels of disparities that can be perceived on a stereoscopic display are far fewer than the number of steps in perspective projection that can be perceived. This is caused by the number of stereoscopic depth layers being a function of the IOD and the horizontal resolution of the display as in Figure 3.7. This is in comparison to perspective projection, where the sample steps are affected by; the distance from the viewer, object size, distance from the line of sight as well as the horizontal and vertical resolution of the screen.

When a 3D image is created, the two views correspond to different view points of the scene. This means that the 2D spatial aliasing artefacts may differ between the two views. This can cause *non-correspondence* aliasing where points cannot be matched correctly between the two views of the image, degrading the stereo perception. An initial study suggested that spatial aliasing, when created by sampling a stereoscopic image, may have a lesser impact on the subjective quality of depth than with the same monoscopic image (Reinhart 1992). A form of *binocular averaging* is thought to take place that attenuates the noise caused by disparate aliasing in the two views. Similar results are found in the stereoscopic volume rendering literature (Hubbold et al. 1997). There, a set of volume rendered images were created at various matching and under sampled data resolutions in monoscopic and stereoscopic environments. Whilst a general preference was found for correctly sampled volumes, stereoscopic images were rated as being higher quality than the equivalent monoscopic images. Despite this, they do however neglect to ascertain whether the non-correspondence aliasing has any objective impact on the performance of tasks.

3.3.2 Smoothing with Depth Perception

Smoothing or blur can affect depth perception in a multitude of ways. When blur is added relative to the depth of features in a scene, then it can be used as a depth cue as discussed in Section 3.1.1. However, blur can be introduced globally to a scene either by display issues such as crosstalk and reconstruction errors such as those caused by non-ideal reconstruction filters. In the HVS, the problem of rays of light not focusing correctly on the retina, called *refractive error*, can also induce a form of smoothing (Goldstein 2002).

In a stereoscopic scene blur can be introduced *monocularly*, in which case only one of the eyes receives an image with some level of blurring and the other eye receives a clear image, or *binocularly* in which case both eyes receive views with the same amount of blurring.

The addition of blur on the stimuli presented to the user can impact the perception of stereopsis and as such has been the focus of a number of studies in the literature. In terms of binocular blur, it has been demonstrated that increasing levels can significantly degrade the results of stereo-acuity tests (Westheimer & McKee 1980, Wilcox et al. 2000, Stigmar 1971). Preliminary results from Stigmar (1971), using a semi-transparent ground glass as the blurring agent, reported that for small angular disparities blur did not significantly impact the perception of depth, with stereo-acuity values being stable at approximately $10''$. However, for larger angular disparities blur was found to cause degraded stereo-acuity values. Extending and formalising the results of this study, Westheimer & McKee (1980) used a Howard-Dolman style 2 Alternative Forced Choice (AFC) test with lenses of increasing optical power placed in front of both eyes. The optical power of the lenses were given in dioptre, a unit of measurement, where a value of 2.0 is equivalent to light rays being brought into focus at 1/2 metres from the eyes. Stereo-acuity was found to degrade more significantly than visual acuity and although contrast was found to lessen as blur increased, a further experiment showed that contrast did not have an effect on the stereo-acuity results. Similar results have been found when Random Dot Stereograms (RDS) have been used (Schmidt 1994). In this case, blur varied from ± 6.0 dioptre in steps of 0.5, it has been found that visual acuity deteriorated regularly with increasing optical blur, where as stereo-acuity fell off at a considerably faster pace with the same value of blur. The results were such that with no blur stereo-acuity could be appreciated at 41", falling to 451" for blur equal to approximately 1.0 dioptre.

Interestingly, the results of Costa et al. (2010) imply that the HVS has some level of

tolerance towards blur with stereo-acuity only increasing when the dioptre increases to 5.0. At the highest dioptre used, 5.0, stereo-acuity increased from $0.1'$ for no blurring to approximately $4.0'$. Furthermore, visual acuity was found to degrade at a similar rate to stereo acuity with participants between the ages of 31 and 45 having increased acuity thresholds when compared to younger participants. The difference between these results and the prior results is the type of test; the experiments of Schmidt (1994) used a RDS stationed at different depths, where as Costa et al. (2010) used a Frisby test that uses physical depth differences to produce disparity. Such results are confirmed by Odell et al. (2009), where the type of stereo-acuity test is found to be an important variable with some tests. Tests based on the RDS performed significantly worse than tests based on real depth presented to the participant, including the Frisby Stereotest and Howard-Dollman test, with the conclusion being that the contours apparent in the real depth tests facilitate the ability to see depths at finer values. Also it was thought that blur can have a larger impact on the smaller dots making up the RDS.

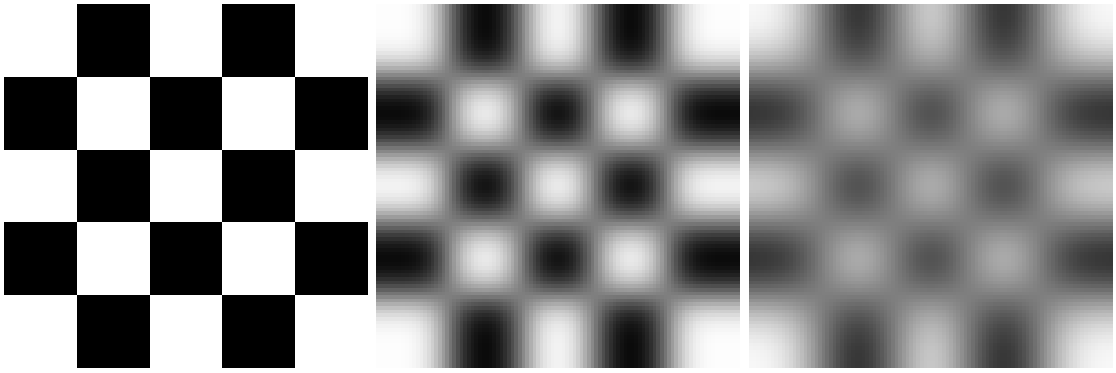
The results from Westheimer suggest that even low levels of dioptre, from 1.5 to 2.0, can have a noticeable difference on the results. This in turn suggests that for fine levels of depth to be perceived, blur should be reduced as a relationship between increased levels of blur and increased stereo-acuity thresholds has been demonstrated. However, not all agree with the results of the prior studies, with the impact seeming to be dependent upon the participant rather than the level of blur (Larson & Bolduc 1991). Using a 4 alternative forced choice test requiring participants to judge which cylinder was nearest Larson demonstrated that out of three participants only one subject experienced the expected increase in stereo-acuity as a function of the blur. Out of the other two participants one had degraded stereo eyesight biasing the results, whist the third experienced almost no loss of stereo-acuity with increasing levels of blur, contrary to the previous results. Whilst it is not clear exactly why the results vary from the previous studies, the differing stereo-acuity tests used and the small user trial may have had some impact on these results.

There are fewer findings on monocular viewing of blur, when only a single view is degraded. The results that have been found suggest that larger levels of blur produce progressively worse values of stereo-acuity than when equivalent levels of blur are being added to both views simultaneously (Westheimer & McKee 1980, Larson & Bolduc 1991). For smaller values of blur added monocularly, no large differences are found from binocular viewing; the HVS seems to be able to compensate for small mismatched values of blur.

From the literature available on blur with stereoscopic displays, there are a few conclusions that can be made. In the cases when blur is added purposefully as a depth cue, an increased perception of depth can be perceived. This is due to the simulation of the real world scenario in which blur gradually increases as the distance from a feature grows. When blur is added binocularly to a view, the majority of results suggest that stereo-acuity should increase as blur does. For monocular blur, low levels may be compensated by the HVS to some degree, however, in high levels significant degradation of stereo-acuity is found. The reasoning as to why blur can degrade stereo-acuity is provided by Wilcox et al. (2000) who suggests that high frequency components in the scene that can aid in the matching of points between two views of a stereo image are attenuated by blur. If corresponding points in a scene are blurred to such an extent that they can no longer be matched, then an implication is that the depth perception of the scene will likely suffer. There is also some evidence that the effect blur has on a stimulus is dependent upon the participant, with either age being a dominant factor or the variation in the participants' acuities (Larson & Bolduc 1991, Costa et al. 2010) .

In terms of volume rendering, there are two stages during the volume rendering pipeline that will smooth or blur the output rendering. If the data is band-limited, so that a low-pass filter is performed on the data to remove high frequencies, smoothing will occur as a result of the fine details being contained in the high frequencies of the data. During the reconstruction stage of the pipeline, in which interpolation filters are used to recreate the continuous function, smoothing depends upon the behaviour of the filter. Non-ideal interpolation filters that do not match the *sinc* filter will induce smoothing to various degrees. This will be binocular in that both views of the image are expected to display the same level of blurring. It is worth noting however the possibility of monocular blurring, as it is likely that reprojection techniques of volume rendering, in which only one of the views is fully rendered, would introduce disparate levels of blur between the two scenes, thereby degrading the depth thresholds more than full stereoscopic volume rendering. Another important factor is that many of the existing studies use an increase in the optical power of lenses placed in front of the eye to increase the level of blur. This is not directly comparable to blurring caused by interpolation errors, yet they do provide some measure of what to expect in blurring in stereoscopic images.

Closer to blurring caused during volume rendering is blur caused by a Gaussian filter being applied to an image. In subjective tests the quality of depth has been tested with



(a) No Blur (b) Slight blur (c) Increased Blur

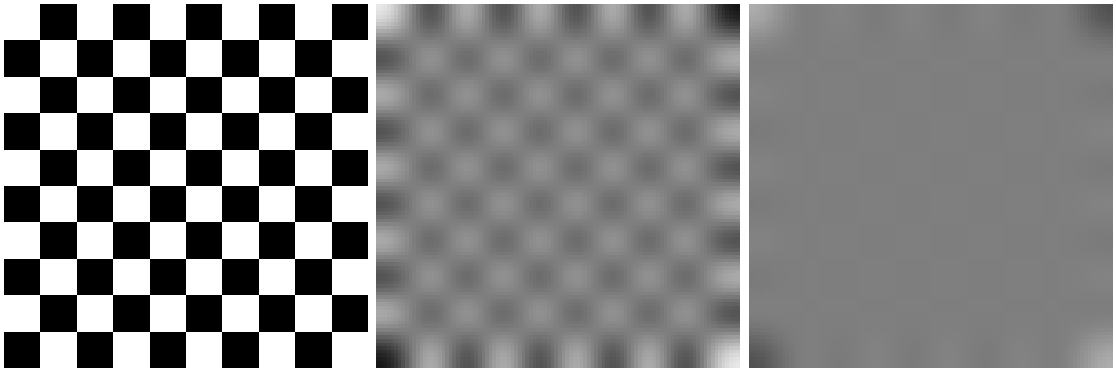
Figure 3.11: A checker board pattern with three different levels of blurring. Fig (a) has no blurring, the Michelson contrast (Michelson 1962) is 1. Fig (b) has been blurred by a Gaussian filter with $\sigma = 5$, the contrast is 0.9167. Fig (c) has been blurred by a Gaussian filter with $\sigma = 8$, the contrast is 0.6357.

different σ parameters of Gaussian blur in binocular stereoscopic images at three different camera base distances (Lambooij 2011). Using the ANOVA method of statistical assessment it was found that the subjective opinion of depth declines as σ increases. This is inline with the objective results of the stereo-acuity tasks. It implies that the results of experiments of blur induced by different dioptre of lenses can be equated to the results of blur caused by low-pass filtering the data sets.

3.3.3 Contrast with Depth Perception

Contrast, the difference of luminance between two objects in the same field of view can be lowered by smoothing the image. When smoothing is introduced into a scene small objects and edges spread into the background, lowering the overall luminance of the background and foreground and decreasing the contrast. In the medical domain, this is a particular problem, if the size of the amount of blur is larger than a feature in the scanned stimulus then, since then it is likely that the feature will have significantly lowered contrast. The impact of blur on contrast is best described in Figure 3.11 where an increasing amount of blur decreases the contrast of a 2D checker board pattern.

While it is difficult to quantify the relationship between blur and contrast, there is a monotonic relationship between the size of the features in the image and the decreased contrast due to blur. Such a relationship can be see in Figure 3.12 where there is a similar pattern to the previous with smaller size and a higher number of checks, and the same



(a) No Blur (b) Slight Blur (c) Increased Blur

Figure 3.12: A more detailed checker board pattern with three different levels of blurring. Fig (a) has no blurring, the Michelson contrast is 1. Fig (b) has been blurred by a Gaussian filter with $\sigma = 5$, the contrast is 0.83399. Fig (c) has been blurred by a Gaussian filter with $\sigma = 8$, the contrast is 0.36759.

amounts of blur causes the contrast to decrease more rapidly.

Depth perception with varying levels of contrast has been well studied, and similar to blur the impact can be segregated into monocular and binocular viewing conditions.

In binocular contrast experiments, the consensus is that decreasing contrast levels make it harder to perceive fine depth levels and as such, increase stereo-acuity thresholds (Legge & Yuanchao 1989, Halpern & Blake 1988, Rohaly & Wilson 1999). A number of studies have determined that a power-law function exists between contrast and stereo-acuity (Rohaly & Wilson 1999, Legge & Yuanchao 1989, Halpern & Blake 1988), suggesting that increasing the contrast continuously should further decrease the stereo-acuity and so provide finer depth perception. Further experiments have found that the relationship eventually stabilises and no further increases in contrast will bring about an improvement in stereo-acuity (Halpern & Blake 1988).

Alternative results have been presented by Ichihara et al. (2007) in experiments in which contrast has been split into two different visual impacts. *Area contrast*, defined as the difference between the average luminance of the surface area of an object and *texture contrast*, the difference between luminance of markings on an object. By using random dot disc patterns with varying levels of contrast to produce texture contrast, and modifying the luminance of the background to create area contrast, it was found that as texture contrast increased the perceived depth of the discs appeared to be closer to the participant in an approximately linear fashion. This follows prior results in which contrast has been used

as a depth cue by using higher levels for nearer objects and lower levels for objects that are to be perceived as further away (O'Shea et al. 1997). Interestingly, Ichihara et al. (2007) found that texture contrast had a lesser impact on depth perception when the area contrast of the image was high, however as the area contrast decreased the texture contrast had a greater impact on the perception of depth. The key difference between theses results and prior is the relationship between contrast and depth. The power-law function of the previous results has not been observed, instead a linear relationship has been reported. It is thought that the main reason for this finding is due to the different methods of testing; the prior experiments used some indirect methods for testing depth, either relying on disparity or time, while the experiments of Ichihara et al. (2007) relied on more direct methods for quantifying depth.

Other aspects of the relationship between depth discrimination and contrast have also been investigated. In the experiments performed in Rohaly & Wilson (1999), it was found that for low contrasts there is a consistent bias favouring uncrossed depths, causing uncrossed stimuli to be perceived with a greater depth and stimuli at crossed depths to be presented with a smaller depth.

The literature we have reviewed so far demonstrates that at low levels of contrast fine depth discrimination is progressively more difficult, yet not all results reflect this. In experiments by Richards & Foley (1974), low values of contrast in photopic conditions are found to improve the accuracy of results for depth ordering in a 3 alternative forced choice (3-AFC) test. This is seemingly contradictory to what might be expected, however the low luminance values are thought to create a stimulus that is better suited towards the analysing mechanisms of the HVS that process large disparities.

In monocular conditions, if the results were to follow the same pattern as with blur, then an expected increase in stereo-acuity would be observed for decreasing levels of contrast. Although in the general sense such results have been demonstrated (Halpern & Blake 1988, Legge & Yuanchao 1989), specific conditions can create exceptions. In particular, the amount of degradation to stereo-acuity is found to vary depending on whether the contrast is low or high when the contrast in the other eye is fixed. In the case when the fixed eye has a high contrast, the results of stereo-acuity are monotonic, degrading as contrast in the variable eye decreases. Halpern & Blake (1988) found that when the variable eye had beyond 15dB of contrast, stereo-acuity became impossible to measure with all stimuli being perceived as collinear in a 3AFC depth-nulling task. However, in

contrasting results, when a low contrast level is presented to the fixed eye, small variances in contrast can actually improve the stereo-acuity thresholds.

In conclusion, the impact of contrast on depth perception implies that higher contrasts increase the ability to perceive finer levels of detail up to a limit of 21dB of contrast. For low values of contrast a perceived bias of depth occurs. Interestingly, low contrast seems to assist in the perception of depth with low resolution stereoscopic displays, the low contrast effectively blurring the scene and introducing a level of anti-aliasing (Jää-Aro & Kjelldahl 1997).

It is worth noting that there is an inverse relationship between blur and contrast, with blurring decreasing the contrast of a scene. This relationship presents challenges in determining the exact impact each one has on perceived depth. As an example, smoothing can be used to simulate blur caused by the limited depth of focus of the eye and contrast can be used to simulate aerial perspective. Yet when they have been applied globally to a scene, fine depth discrimination suffers.

In relation to the impact on depth perception in volume rendering, the introduction of blur can decrease the contrast of the rendering. The type of rendering algorithm can also make a difference upon the expected result; MIP and X-Ray renderings can produce images with higher degrees of contrast than full volume rendering (Boucheny et al. 2009).

3.3.4 Conclusions

In this chapter an overview has been given of the different methods that can be used to perceive depth in the HVS. These include pictorial, oculomotor and binocular depth cues. Although each cue can aid depth estimations they are not all effective in the same range and it is still an active research area into how the cues are combined by the HVS.

This chapter has also reviewed 3D displays and the different hardware designs that can be used to display 3D images. The issues involved with displaying 3D images including camera geometry, crosstalk, aliasing and smoothing have also been given with details on how these impact how depth is perceived.

Chapter 4

Investigating Depth Perception with Stereoscopic Volume Rendering

This chapter is based on the paper 'Investigating Depth Perception with Stereoscopic Volume Rendering' (Roberts 2014). It is an experimental study evaluating the effect of volume resolution on the depth thresholds obtained via a stereo-acuity test.

4.1 Introduction

The accurate representation of depth in volume rendered scenes is a key requirement in order to ensure correct diagnostic evaluations. With the increase in availability of 3D displays over the last few years there is a need to understand how stereoscopic depth is presented with direct volume rendering. Initial work in this area has primarily focused on how the spatial ordering of stimuli and the ability to determine structure from motion is impacted by differing levels of assigned opacity. The rate at which the volume is sampled and the choice of interpolation scheme used to reconstruct the discrete data set can cause errors known as aliasing as well as post-aliasing and smoothing. How these impact the perception of depth in stereoscopic DVR is not fully understood.

In this chapter, we evaluate a DVR algorithm, using three volume resolutions, against a non-DVR reference algorithm in terms of the depth threshold, δ, obtained when using a stereo-acuity test. We performed two experimental studies, the first displaying just the stimulus for the stereo-acuity test and the second displaying a window in front of the

stimulus to reduce size as a factor in the tests.

The analysis of the results suggest that the depth threshold is not significantly affected by the resolution of the volume. That is, users were still able to perceive fine depths regardless of the volume resolution or the use of the reference algorithm. However, when the response time of a participant to make a depth judgement was analysed, it was found that response time generally increased with higher volume resolutions.

The implication of the results is that volume resolution does not have a strong relation to the depth threshold that can be achieved when using DVR. Secondly, if minimal response time is important, then lower volume resolutions should be preferred.

4.1.1 Depth Perception with Stereoscopic DVR

Stereoscopic volume rendering is a relatively recent research area and as such there are a dearth of formalised studies. The earliest investigation into the impact of stereopsis was performed by Hubbold et al. (1998) targeting the impact of different levels of transparency with volume rendering. In the experiment, participants were required to determine the depth ordering of a voxelised sphere in relation to two neighbouring spheres at different voxel disparities. Spheres were used as these were thought to be indicative of realistic data sets. It was found that whilst the use of the stereoscopic display did generally improve the results, this was not unanimous nor consistent over all voxel disparities; in particular smaller uncrossed disparities found to produce conflicting results with shading. In a somewhat similar study, the perceived relative depth of cylinders contained within a computed-tomography data set had to be determined aided by various rotational modes of the cylinders (Martinez Escobar et al. 2013). Although no clear benefit of the stereoscopic display of the data was found, it seems that the colours used for the cylinders red, blue and green are likely to have induced chromostereopsis and therefore introduce an unexpected variable into the tests.

In studying the recovery of structure from motion of an object (Kinetic Depth Effect) Kersten (2006) and Agus et al. (2009) have demonstrated that determining the rotation of volume rendered cylinders can be significantly improved by the use of stereoscopic displays. Despite the fact, that in both cases the cylinders contained Perlin noise and therefore produced highly ambiguous renderings. This suggests that the spatial understanding of interactive volume renderings may be greatly improved by stereoscopic displays even for data that is ambiguous, or noisy, for example ultrasound data. Utilising a similar

experiment, it was found that the opacity, controlled by the transfer function, can also have a significant impact on the ability to understand the spatial relationships of features within a volume (Boucheny et al. 2009).

4.1.2 Aliasing with Depth Perception

A few studies have evaluated aliasing with depth perception. With experiments performed with DVR Hubbold et al. (1997) found that there was a preference for correctly sampled volumes. However, it was also reported that the HVS can ignore some aliasing artefacts. This contrasts with objective studies on depth aliasing caused by limited depth planes as explored by Jää-Aro & Kjelldahl (1997). There, requiring users to assess the depth of a shape given the depths of two other shapes in the scene, it was found that at low resolutions, pixels within the images did not align causing misleading shapes and ghosting, resulting in lower accuracy rates. It was concluded that stereoscopic display of the data was less crucial than anti-aliasing of the images. This demonstrates that aliasing can have a decidedly negative impact on the perception of depth, negating in some cases the use of stereoscopic displays.

More detailed experiments have shown that sampling in stereoscopic displays can cause inaccuracies and inconsistencies in the position, size and disparities up to half a pixel in size (Pfautz 2000). These are formed due to the mismatch between the number of perspective steps that can be displayed versus the more limited number of disparity steps.

Although investigations have looked into aliasing (Hubbold et al. 1997) and more generally into how depth can be appreciated in volume rendering (Martinez Escobar et al. 2013) none attempted to determine how fine depth perception is impacted by aliasing and reconstruction errors within stereoscopic volume rendering.

4.2 Experiment 1

In order to investigate how fine depth is impacted by volume resolution the Howard-Dollman *stereo-acuity* test (Howard 1919) has been used with the adaptive Parameter Estimation by Sequential Testing (PEST) methodology (Taylor & Creelman 1967) in order to rapidly converge on the minimum depth for a series of renderings. In total we test four conditions, 3 different volume resolutions and 1 non-DVR reference rendering that does not require sampling. Details regarding the methodology and stimuli are described in the

following sections.

4.2.1 Stereo-acuity

In order to assess how aliasing and reconstruction errors can impact fine depth perception it is necessary to use a test that finds thresholds for the smallest differences in depths that can be accurately determined between stimuli. Such tests are known as *stereo-acuity* tests. The Howard-Dolman stereo-acuity test is one such example (Howard 1919).

In the original Howard-Dolman stereo-acuity test, two wooden 1cm diameter cylinders are stationed 6cm apart from each other. During each experiment one of the cylinders, designated the test, is moved at some distance away from the other cylinder designated the control with the initial starting depths being 6000mm and 6030mm from the participant respectively. At no point are the two cylinders at the same depth. The participant is then asked to state which of the cylinders is nearer to themselves.

For the purpose of this experiment a modified version of the Howard-Dolman test was used. The cylinders were rendered on a 3D display, specifically a True3Di 24 inch display that is capable of projecting a full HD resolution of 1920×1200 with participants being seated 700mm away from the display. Four sets of cylinders were rendered, three were rendered using a DVR algorithm with volume resolutions of 512^3, $512^2 \times 256$ and $512^2 \times 128$. The fourth set were generated using the non-DVR reference algorithm that used a simplified form of the volume rendering integral as discussed in Section 4.2.3. Therefore, in total there were four different trials. The setup of the experiment is shown in Figure 4.1.

The depths the cylinders are moved by are called steps, and the sizes of these vary depending on the number of correct judgements made by a participant. This prevents the minimum depth threshold from being skipped over and potentially decreases the time taken to find the smallest threshold value. For each judgement a rendering of two cylinders is displayed, the user must then state via pressing the corresponding left or right arrow key on the keyboard which of the two cylinders is closest to them. In order to remove learning effects from the trial, the order of which cylinder is closest is randomised for each user when the trial is started. When the user has made a response, the screen is blanked with a waiting image which is shown for two seconds to remove residual imprints of the last image that was judged. Details regarding the adaptive step procedure used are given in the next section.

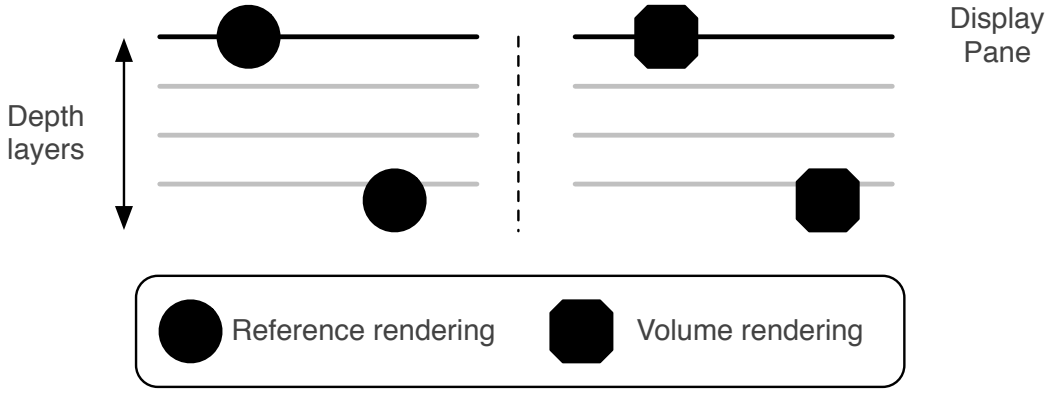


Figure 4.1: Top down view of the stimulus in the experiment. The positions of the cylinders correspond to the initial starting positions.

4.2.2 Parameter Estimation by Sequential Testing

In order to find the smallest depth threshold that can be perceived between stimuli the following methodology is used. After a user makes a set number of judgements at a depth level, the depth between the two rendered cylinders will change. For each series of judgements at the same depth, the number of correct judgements and the total number of judgements are kept. A test is run after each trial that defines a lower and upper bound for the permissible number of correct choices. If the number of correct choices is greater than or equal to the upper bound then the current depth level is too large and a smaller depth level is used. If it is on or below the lower bound, then the current testing level is considered too difficult and a larger depth difference is used. The change in direction from an increase in depth to a decrease, or vice-versa, is known as a reversal.

To find values for step sizes for each judgement as well as starting and stopping conditions the Parameter Estimation by Sequential Testing (PEST) methodology has been followed (Taylor & Creelman 1967). This is an adaptive procedure that can be used to rapidly find thresholds for an independent variable. To generate the step size for each judgement a set of four rules are used that depend on the direction of the last judgement, the number of judgements in the same direction and conditions that lead up to the last reversal of direction. The rules are defined as:

1. If a reversal takes place, the step size of the next judgement is half the previous size in the opposite direction.

2. For a second judgement in the same direction the step size is the same as the first

step size in that direction.

3. For a third judgement in the same direction the step size depends on the events that caused the last reversal:

   (a) If the most recent reversal resulted from a doubling of the step size then the size does not change.

   (b) If the last reversal was not due to a doubling then the step size for the next judgement is doubled. The purpose of this rule is to prevent the same sequence of step sizes from being repeated.

4. For a fourth judgement in the same direction the step size is double the previous step size.

The minimal step size, used to determine when to end the trial, can be calculated from the viewing distance which is $700mm$, the IOD which is $65mm$ and the width of a single pixel, equal to $0.27mm$, with the 3D display used. The IOD used is the typical value given for adults (Holliman 2005). Although no children participated in the following experiment, if they were to it would be necessary to adjust the IOD accordingly to ensure the same depths are perceived. By substituting these values into the equation for perceived crossed disparity (Holliman 2005):

$$P = \frac{z}{\frac{IOD}{|d|} + 1} = \frac{700}{\frac{65}{0.27} + 1} \tag{4.2.1}$$

the smallest perceived depth is $2.9mm$, equivalent to 79.4 seconds of arc.

The PEST methodology advises a starting value that is easy for users to be able to discriminate. At $700mm$ a lower bound for the maximum perceived depth that can be displayed before images can no longer be fused is $60mm$ (Jones et al. 2001). This was used as the initial starting depth of the test cylinder.

4.2.3 Rendering Algorithms

Two rendering algorithms were used in the experiment. In order to produce an image free of sampling and interpolation a reference algorithm has been used based on earlier work from Hsu & Chelberg (1993) and Max (1995). The second algorithm is a standard volume rendering algorithm using the ray-casting method.

It has been decided not to use surface shading in the renderings. Although shading is a depth cue and can help the understanding of the positioning of objects (Goldstein 2002) it

has been found to conflict with fine depth perception when stereoscopic volume rendering is used (Hubbold et al. 1997). Further, when ultrasound data is used that contains noise, shading is not employed due to the difficulty in obtaining accurate gradients.

For all renderings, perspective projection has been used. Although this introduces a depth cue, if stereoscopic images are presented without it an effect known as *perceptual zooming* is introduced into the renderings. This causes objects that are further away to appear larger than objects that are closer. It is possible that such an effect may be a cause of the negative results when comparing stereoscopic against monoscopic volume renderings in previous experiments (Ropinski et al. 2006).

Reference

The full volume rendering integral is defined as (Max 1995):

$$I(D) = I_0 T(D) + \int_0^D C(s)\tau(s)e^{(-\int_S^D \tau(t)dt)}ds \tag{4.2.2}$$

where $I(D)$ is the intensity of the ray to position D, I_0 is the background intensity, $T(D)$ is the transparency through the volume medium to position D, $C(s)$ is the colour at position s and $\tau(s)$ is the extinction coefficient that defines the rate at which light is occluded.

In practice the full volume rendering integral cannot be computed analytically, therefore some simplifications are required. The following simplifications are derived from the work of Max (1995). By assuming that the colour parameter, $C(s)$ is constant along a ray, we get:

$$C\int_0^D \frac{d}{ds}e^{(-\int_S^D \tau(t)dt)}ds \tag{4.2.3}$$

which is equivalent to:

$$C(1.0 - e^{(-\int_S^D \tau(t)dt)}) \tag{4.2.4}$$

Putting this back into the full volume rendering integral we have:

$$I(D) = I_0 e^{(-\int_S^D \tau(t)dt)} + C(1.0 - e^{(-\int_S^D \tau(t)dt)}) \tag{4.2.5}$$

The colour C is composited on top of the background I_0 with the transparency $e^{(-\int_S^D \tau(t)dt)}$. By using a background with intensity 0, or black, the first term of the equation can be removed producing:

$$I(D) = C(1.0 - e^{(-\int_S^D \tau(t)dt)}) \tag{4.2.6}$$

Now, the only integral left is in the power function that defines the rate of absorption within the volume medium. The integral $-\int_S^D \tau(t)$ is the optical depth from the position

S through to the eye position at D. If τ is not dependant upon the position t then a further simplification can be made so that:

$$e^{-\int_S^D \tau(t)dt} \equiv e^{-\tau D} \tag{4.2.7}$$

If $-\tau$ is replaced by $-\mu$ and D with x then this equation becomes the Beer-Lambert law that relates the absorption of light to the medium the light is travelling through.

$$e^{-\mu x} \tag{4.2.8}$$

In this case μ is the attenuation coefficient and x is the object thickness. The equation then becomes:

$$I(D) = C(1.0 - e^{-\tau D}) \tag{4.2.9}$$

This equation is then straight forward to compute analytically by iterating through each pixel and firing a ray into the scene. Note that this assumes that the medium is constant and has a constant transfer function. It restricts the features in the volume to uniform colour with opacity that is defined by the density of the feature. Such restrictions allow for a volume to be defined with no sampling or interpolation required. To render the cylinders, the equation is calculated for the intersection points between a ray and the cylinders.

In a series of investigations a similar method has been used to create X-Ray like renderings of complex geometric shapes including simulated models and parts of the human body including knee-joints (Hsu & Chelberg 1993). It was also used for the creation of X-ray like images for the detection of abnormalities in simulated breast tissue in stereoscopic environments (Hsu et al. 1993, Hsu & Babbs 1993, Chelbergy et al. 1994).

Direct Volume Rendering

A in-house software-based stereoscopic ray-casting algorithm has been used, the details of which can be found in Appendix B. All calculations in the renderer are performed with double-precision floating-point formats. In order to ensure accurate reproduction of the volume data, the sample rate along each ray has been set above twice the highest volume resolution used, in this case ($512 \times 512 \times 512$). The trilinear interpolation scheme has been used as the reconstruction filter. A 1D transfer function with an opacity that increases linearly from from 0 through to 0.3 and a constant green value of 0.35 has been used.

Stimuli

The reference rendering algorithm and the ray-casting algorithm create the renderings based on the same data source. A pair of cylinders are used that have a radius of $12.26mm$ when projected onto the screen plane and subtend an angle of $1.003°$ horizontally. The length of the cylinders are $259.2mm$ when projected onto the screen plane, equivalent to subtending an angle of $20.978°$ vertically.

The reference rendering uses the geometric representation of the cylinders and the simplified volume rendering integral as detailed in Section 4.2.3.

The standard volume rendering ray-casting algorithm used in the experiment requires the use of a method to *voxelize* the cylinders. A filter based voxelisation system based on the work of Sramek & Kaufman (1998) has been developed for this experiment to produce the volume data sets.

Three different volumetric resolutions have been chosen for this experiment, $(512 \times 512 \times 512)$, $(512 \times 512 \times 256)$ and $(512 \times 512 \times 128)$ with each value being an 8-bit integer. The anisotropic data sets, with the samples in the z direction of 512, 256 and 128 reflect typical Computed Tomography data sets where the resolution in the z-axis is less than the resolution along the slice (x, y). The aspect ratio of the data set sizes are $(1, 1, 1)$, $(1, 1, 2)$ and $(1, 1, 4)$ for the three data set sizes respectively. Figure 4.2 shows an example rendering of the cylinders displayed in the experiment.

4.2.4 Participants

16 participants aged between 22-57 were recruited to take part in the experiment, with 7 women and 9 men. Prior to the trial all participants were screened for vision by using the Bailey and Lovie Snellen chart and for stereo-vision using the Titmus fly stereo test. Only those participants with 20/30 vision, capable of perceiving letters $9mm$ high at $6m$ away, and have successfully passed the Titmus fly stereo test were asked to proceed.

Each participant filled in a pre-experiment questionnaire to ensure that all instructions were understood and that any questions they may have had about the experiment were answered. The experiment was approved by the ethics committee of the School of Engineering and Computing Sciences, Durham University.

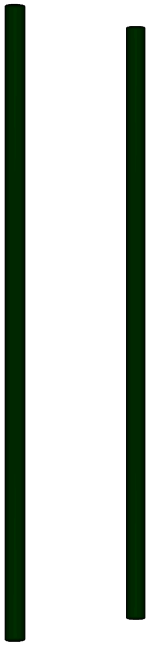


Figure 4.2: Rendering of the reference stimulus used in Experiment 1 with white background.

4.2.5 Equipment and Viewing Conditions

A True3Di 24″ HD-SDI Monitor was used for displaying the graph images in both the stereoscopic and monoscopic test setups. The display provides a WUXGA resolution of 1920×1200 pixels for each eye with a refresh rate of $60Hz$. In order to view the 3D images, linear polarized glasses were required to be worn.

Participants were placed at a distance of $60cm$ from the display. The experiment took place in a darkened room with the light levels kept consistent throughout the experiments.

4.2.6 Procedure

Prior to the experiment each participant took part in a pre-trial where depth thresholds were not recorded in order to familiarise themselves with the type of experiment. There were four trials in total, one for the non-DVR reference algorithm and one for each of the volume resolutions tested. Between each trial there was a short break to reduce participant fatigue. A Latin Square design was used to determine the order of the trials to minimise learning effects (Montgomery 2009). In each trial the smallest depth threshold achieved by the participant was recorded as well as the time taken to make a decision for each of

Table 4.1: Mean Depths Perceived (pixel disparity)

Reference	$512^2 \times 128$	$512^2 \times 256$	512^3
1.094	1	1.094	1.25

the depths tested.

4.2.7 Hypothesis

An expected result of this study is that the non-DVR reference algorithm will lead to the smallest depth thresholds from the experiment. Further, that out of the three volume resolutions tested, the highest resolution of $(512 \times 512 \times 512)$ will produce depth thresholds similar to the non-DVR reference algorithm.

4.3 Results

4.3.1 Depth

Via the use of the PEST methodology a single depth threshold for each trial for each participant was retrieved. The mean depth thresholds as pixel disparity for each of the trials are shown in Table 4.1.

Shapiro-Wilk's test determined that the depth thresholds were normally distrubuted. A one-way ANOVA statistical test was performed with the result being $F(3, 45) = 1.754$, $p = 0.170$. Therefore, there were no significant differences between the depth thresholds that could be perceived between each of the trial conditions.

4.3.2 Response Latencies

During each trial the response time a user took to make a judgement was recorded. Table 4.2 shows the mean response times for each of the pixel disparities that were tested. Response times were only recorded when a correct response was given (Pfautz 2000). Due to the use of an adaptive procedure not all participants were tested at each of the depth levels, therefore to allow for an accurate comparison only the response times at individual depths were analysed.

A one-way analysis of variance indicates that at 1 pixel disparity the choice of rendering data set had a weakly significant impact on the response time taken to make a correct

Table 4.2: Response Latencies (seconds)

Disaprity	Depth(mm)	Reference	$512^2 \times 128$	$512^2 \times 256$	512^3
1px	2.896	2.794	1.901	2.193	2.326
2px	5.767	2.220	-	1.758	-
3px	8.616	2.184	-	1.853	-
4px	11.440	1.777	1.226	1.352	1.358
5px	14.243	-	-	-	-
6px	17.022	-	-	-	-
7px	19.779	-	-	-	-
8px	22.513	1.211	1.124	1.119	1.162
9px	25.226	-	-	-	-
10px	27.916	-	-	-	-
11px	30.587	-	-	-	-
12px	33.236	1.228	0.99	0.954	1.082
13px	35.863	-	-	-	-
14px	38.470	-	-	-	-
15px	41.057	-	-	-	-
16px	43.624	1.078	1.372	1.134	1.05
17px	46.170	-	-	-	-
18px	48.697	1.466	1.387	1.229	1.142

choice with $F(3, 45) = 1.638$, $p = 0.083$. Using a post-hoc Tukey HSD, adjusted to take into account unbalanced sample sizes, it was found at 1 pixel of disparity the reference rendering was weakly significantly different from the rendering using the $(512 \times 512 \times 128)$ data set size with $p = 0.059$.

4.3.3 Participant Comments

Although as part of the formal user trial subjective comments were not requested, informal comments were recorded. It was noted that for larger depths differences although the depth difference made it obvious which cylinder was in front of the other, a number of participants stated that their strategy for the larger depths was to determine which cylinder was larger either length or width wise. Further, one participant commented that

in stimuli that showed particularly blurry content for smaller depth disparities the strategy they used was to locate the tops and bottoms of the cylinders to determine depth from their position.

4.3.4 Discussion

Contrary to the hypothesis, the non-DVR reference algorithm was not found to produce the smallest depth thresholds with there being no statistical difference between the reference rendering and the different volume resolutions tested. From the results obtained it would seem that slice thickness and resulting aliasing in the rendered image have no impact on the perceived depth, despite prior results to suggest otherwise (Hubbold et al. 1997, Jää-Aro & Kjelldahl 1997, Pfautz 2000). As became apparent after the experiment, participants noted that the changing in size of the cylinders provided a convenient method of determining relative depth of the cylinders and may account for the lack of differences between rendering types, despite participants noting increased blurriness and striping.

Regarding response time, at smaller disparities the response times increased. This may indicate an increased level of difficultly in determining fine depth differences. Further, a weak significance was found between the reference rendering and the $512^2 \times 128$ dataset at the smallest depth tested. This suggests that the differences between the renderings may be more detectable at small depths.

In order to reduce the impact of perspective on the results a second experiment was developed, as discussed in the following section.

4.4 Experiment 2

4.4.1 Details

From the informal comments given by the participants it became apparent that despite orthoscopic perspective projection being used, the difference in lengths between cylinders was providing a trivial method for participants to determine which cylinder was in front of the other. However, using orthographic projection leads to *perceptual zooming* an artefact causing closer objects to be perceived as being smaller than objects further away. To ensure that binocular vision is the main cue used to determine the differences in depth for each of the cylinders a frame is rendered in front of the cylinders. The frame covers both ends of the cylinders and so removes the ability to use length as a cue. Figure 4.3 shows

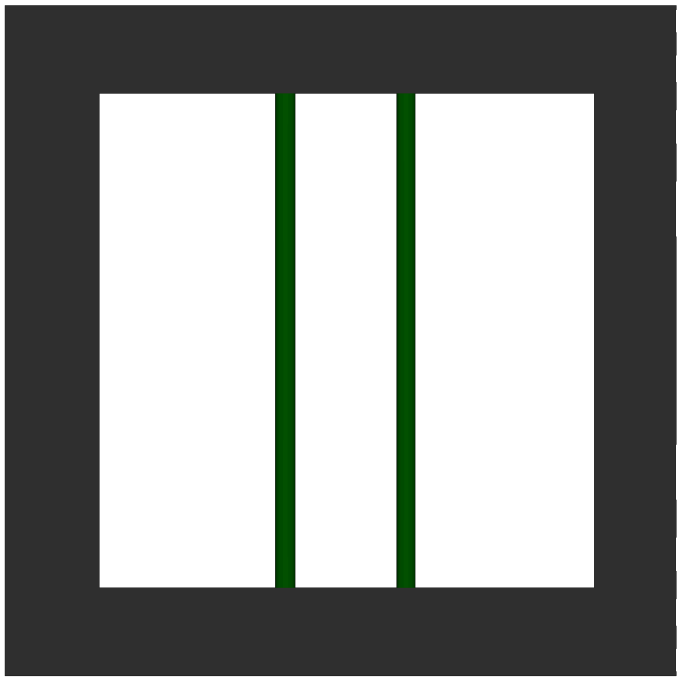


Figure 4.3: Example rendering of the reference test stimulus used in Experiment 2 with the frame rendered in front of the cylinders to reduce perspective cues.

renderings of the cylinders with the included frame.

4.4.2 Procedure

The setup and design of the experiment follows that of the previous experiment. There were four conditions in total, 3 different volume resolutions and 1 non-DVR reference rendering.

4.4.3 Participants

16 participants aged between 22-57 were recruited for this experiment. All participants were tested with the Bailey and Lovie Snellen chart to ensure 20/30 vision and the Titmus test for stereo-acuity. A pre-questionnaire was used to ensure that all instructions were understood. All participants were novices in that they had no professional experience of volume rendering.The experiment was approved by the ethics committee of the School of Engineering and Computing Sciences, Durham University.

Table 4.3: Mean Depths Perceived (pixel disparity)

Reference	$512^2 \times 128$	$512^2 \times 256$	512^3
2.7	3.733	3.667	4.567

4.4.4 Results

Depth

Table 4.3 shows the mean depth thresholds for each volume resolution and reference rendering. As assessed by Shapiro-Wilk's test the depth thresholds were not normally distributed ($p < 0.001$). Therefore, the non-parametric Friedman test was used. The depth thresholds increased from reference (Mdn=2.700) to 256 (Mdn=3.667), to 128 (Mdn=3.733), to 512 (Mdn=4.566) but the differences were not statistically significant, $\chi^2(3) = 3.911$, $p = 0.271$.

Response Latencies

During the experiment response times were recorded for each correct step judgement. Table 4.4 shows the mean response times for each of the pixel disparities. Due to the use of the PEST methodology not all participants viewed the stimuli at each depth.

At 18, 16 and 12 pixels differences were found, the repeated-measures ANOVA and the Bonferroni test were used for post-hoc analysis. All latencies were normally distributed ($p > 0.05$). In each case sphericity had been violated with $p < 0.001$ therefore the Greenhouse-Geisser correction was applied. At 18 pixels the results were $F(1.331, 25.284) = 0.667$, $p = 0.002$ with increases found between the reference rendering and the $(512^2 \times 128)$, $(512^2 \times 256)$ and (512^3) volume resolutions, with $p = 0.024, 0.047, 0.006$ respectively. At 16 pixels the results were $F(1.214, 18.211) = 12.123$, $p = 0.002$. Increases were found between the reference rendering and the (512^3) volume resolution with $p = 0.017$. Further differences were found between the $(512^2 \times 128)$ and (512^3) volume resolutions with $p = 0.010$ as well as the $(512^2 \times 256)$ and (512^3) volume resolutions with $p = 0.020$. At 12 pixels the results were $F(1.952, 42.954) = 9.194$. Increases were found between the volume resolutions of $(512^2 \times 128)$ and (512^3) with $p = 0.004$ and the $(512^2 \times 256)$ and (512^3) volume resolutions with $p = 0.006$.

The latencies were then aggregated across disparities. The data was not normally

Table 4.4: Response Latencies (seconds)

Disparity	Depth(mm)	Reference	$512^2 \times 128$	$512^2 \times 256$	512^3
1	2.896	4.273	3.073	3.358	4.790
2	5.767	2.687	2.897	3.457	4.201
3	8.616	4.915	3.819	2.294	2.181
4	11.441	2.252	2.412	2.11	2.785
5	14.243	-	3.416	2.291	1.523
6	17.022	4.976	1.973	1.937	1.694
7	19.779	-	-	-	-
8	22.513	2.760	2.117	1.783	1.964
9	25.226	7.037	0.9928	-	-
10	27.917	3.074	-	1.078	1.601
11	30.587	8.109	3.469	-	-
12	33.236	3.06	1.881	1.655	1.847
13	35.863	-	-	-	-
14	38.470	3.120	4.187	2.111	2.686
15	41.057	3.061	2.164	-	-
16	43.624	2.684	1.963	1.618	2.961
17	46.170	2.365	2.316	3.564	-
18	48.697	4.476	3.276	2.933	3.583
Aggregate		3.539	2.639	2.285	3.101

distributed, $p < 0.001$. Therefore the Friedman test was used. Latencies were significantly different between renderings with $\chi^2(3) = 21.468$, $p < 0.001$. Post-hoc analysis revealed differences between the reference (Mdn=1.937) and ($512^2 \times 128$) volume resolution (Mdn=1.688) $p < 0.001$. Further differences were found between the reference and ($512^2 \times 256$) volume resolution (Mdn=1.641), $p = 0.001$.

Between Analysis

To analyse the effect of the window on the depth thresholds we perform an in-between comparison of the combined results from the two experiments. Shapiro-Wilk's test determined that the depth thresholds were not normally distributed, $p < 0.001$, therefore

the non-parametric Kruskal-Wallis H test was run to determine if there were differences in depth thresholds between the two experiments. Median depth thresholds were statistically significantly different between groups, $\chi^2(1) = 16.891$, $p < 0.001$. The median depth thresholds for both experiments were 1.0, however Experiment 1 had a mean depth threshold of 1.109 and Experiment 2 had a mean depth threshold of 3.667.

The response times were also compared between Experiment 1 and Experiment 2. Shapiro-Wilk's test determined that the response times were not normally distributed, $p < 0.001$, therefore the non-parametric Kruskal-Wallis H test was run to determine if there were differences in response times between the two experiments. Median response times were statistically significant different between groups, $\chi^2(1) = 12.012$, $p = 0.001$. The median response time for Experiment 1 was 2.055 seconds and the median response time for Experiment 2 was 2.469 seconds.

4.4.5 Discussion

The results suggest that volume resolution has little impact on depth perception. This would seem contrary to the effects of sampling in a static stereo scene given by Pfautz (2000) as well as the subjective responses of under sampling a volume rendered scene from Hubbold et al. (1997).

From further evaluation of the results and of the stereo settings used for the experiment a potential explanation can be given for the lack of statistically relevant difference between the data resolutions and the reference rendering. When reconstructing a continuous signal from a discrete signal using linear interpolation the largest error can be expected midway between existing sample points. In the case of volume rendering this would be between two existing voxel samples. Assuming the data set resolution of $(512 \times 512 \times 128)$ and that a voxel lies on the screen plane, the difference in pixel disparity to the next voxel in the z direction is 0.484 pixels. With the display settings used this is equivalent to 38.354 seconds of arc in depth. With the trilinear interpolation scheme used, the largest error should then be expected at approximately 19.227 seconds of arc. Currently, the smallest depth that is tested is a single pixel equal to 79.387 seconds of arc, this is larger than where the errors should be expected. Therefore, there is the potential for the artefacts to be more distinguishable at smaller depths than those that are considered for this experiment. This can be fulfilled by either using sub-pixel disparities or repositioning participant and virtual camera at a distance further away from the screen plane.

Compared to the depth scores, the response times recorded produced statistically more significant results that suggests an interesting relationship between data set resolution size and the amount of depth on response times. The results at 18, 16 and 12 pixels of disparity seem to suggest that the reference and high resolution volumes require a greater duration of time in order to determine a correct answer than for lower resolution volumes. When all the response times for each pixel disparity are combined, in general the reference rendering takes a greater duration than lower resolution volumes. This may indicate that at higher disparities, smooth data with few discernible edges can impact the ability for a user to correctly perceive the 3D scene. A possible reason is that when fusing stereoscopic images the HVS may use zero-crossings to match pairs of images (Anderson & Nakayama 1994). Zero-crossings correspond to luminance edges that can relate to object boundaries. For lower resolution volumes there are a higher number of luminance edges that may assist in determining correct depths and reduce response latency.

Although the results seem to suggest higher resolution volumes require a longer time to accurately determine the depth, there are two factors that should be considered. Firstly, the timings are being studied as a secondary factor, in particular due to the use of the PEST methodology it is possible that each participant may have performed a different number of judgements at different pixel disparities. Further, several pixel disparities were not tested due to these being jumped over when trying to find a depth threshold.

Significant results were found when the depth thresholds were compared between Experiment 1 and Experiment 2. The smaller depth threshold achieved from the first experiment confirm our suggestion that participants were using the different length of the cylinders to aid in making depth judgements. The longer response times from Experiment 2 may also suggest that participants found the task more difficult when they had to rely principally on the binocular depth cue.

4.5 Conclusions

We have presented an investigation into the effect of volume resolution on fine depth perception with stereoscopy extending the work of Hubbold et al. (1997). The results suggest that response latency and not depth is more greatly impacted.

The study itself has influenced the design of later experiments documented in this thesis. The use of abstract datasets, such as the cylinders, were inspired from psychology, however, their realism is low and therefore later experiments use datasets with character-

istics similar to real datasets. Further, in real-life, it is necessary to gain an understanding of an entire dataset, rather than just the raw depths between two features. Therefore, later experiments will use tasks that require a global understanding of the scene.

The study presented has limitations, the first of which is the abstract nature of the cylinders that do not correspond to real life data sets. The second limitation is the method of voxelisation used to generate the cylinder data sets, the implementation used is highly specific for the cylinders and not generalisable to other types of geometry.

Chapter 5

An Evaluation of Reconstruction Filters for a Path-Searching Task in 3D

This chapter is based on work published in 'An Evaluation of Reconstruction Filters for a Path-Searching Task in 3D' (Roberts et al. 2014). It is an experimental study evaluating the effect the choice of reconstruction filter has on a task requiring spatial understanding in 2D and 3D. We find that the 3D condition is more accurate than 2D and that the reconstruction filter can have an impact on task accuracy.

5.1 Introduction

In this chapter, we evaluate three different reconstruction filters used in DVR in terms of their accuracy and response times when used to produce images for a path-tracing task. We compare trilinear, Catmull-Rom and interpolating B-spline reconstruction schemes in stereoscopic and monoscopic environments. Despite research into the analytical behaviour (Marschner & Lobb 1994, Möller et al. 1996), image quality and subjective opinion of reconstruction artifacts (Mitchell & Netravali 1988), it is not well understood how the choice of reconstruction filter may impact task performance or whether task accuracy relates to the analytic quality of the reconstruction filters tested. Further we seek to determine the possible benefits that stereoscopic displays may have against monoscopic displays when used with DVR for a spatial search task.

5.1.1 Stereoscopic Volume Rendering

Whilst there is evidence to suggest that stereoscopic volume rendering can improve the results of tasks, this is not conclusive neither is it consistent even between similar tasks.

When comparing the depth of two vessels in angiography data sets, Ropinski reported that participants who were inexperienced with stereoscopic displays produced increased error rates and response times when compared to monoscopic depth cues (Ropinski et al. 2006). Likewise, for a similar task performed by novices and experts Kersten reports that enhanced monoscopic depth cues were capable of producing better results than stereo (Kersten-Oertel et al. 2014). In Cho et al. (2014) in contrast, for depth discrimination tasks using simulated angiography-like data sets stereoscopic displays proved more effective in most cases against 2D displays. Recently, Abhari et al. (2015) investigated methods to enhance MR angiography images by combining contour enhancement with stereopsis for DVR. Using a relative depth task it was found that whilst stereopsis improved accuracy, it was most effective when combined with contour enhancement.

Purely absorptive volume rendering has also been evaluated when combined with the kinetic depth effect in stereoscopic and monoscopic conditions (Kersten et al. 2006). Stereoscopic displays were found to improve the ability of users to determine which direction a volume rendered cylinder was rotating. Similar results for the same task were found in a later study with multi-view auto-stereoscopic displays (Agus et al. 2009).

Further conflicting results have been reported when participants have been required to determine the relative depth of volume rendered translucent cylinders. An initial study reported that stereo display of the cylinders is beneficial for this type of task (Cho et al. 2012). However, in a similar study with cylinders combined in a computed tomography data set, stereoscopy was found to have no significant effect for correct depth discrimination (Martinez Escobar et al. 2013).

Even within the same study the results of tasks for stereoscopic volume rendering are not consistent. In an early investigation of volume rendering with stereoscopy, Hubbold found that when users were required to determine which sphere from a set of three was nearest, stereo only improved the results when the spheres were contained within a semi-transparent shell (Hubbold et al. 1997). A recent study (Laha et al. 2012) evaluating a range of quantitative tasks for volume rendering with different immersive display setups found that complex search tasks benefited from higher levels of immersion. Yet when slicing the data sets a decrease in performance was found when stereoscopic displays were

used. The follow-up study found that different display modalities benefit different volume rendering styles (Laha et al. 2014).

A possible reason for the range of results is the variation in the parameters used for rendering the images. These include data set resolution, transfer functions as well as stereo settings. In this evaluation we seek to maintain settings between trials with only the selected interpolation filter changing to determine how this parameter may effect the results of a spatial task in stereoscopic and monoscopic environments.

5.1.2 Reconstruction in Volume Rendering

Recapping from Chapter 2, volumetric data, produced either from medical scanners or via a process such as voxelisation, is represented in the form of an array of sample points on a 3D grid lattice. To render volume data it is necessary to reconstruct values at points x between the grid nodes k. This can be written as (Thévenaz et al. 2000*b*):

$$f(x) = \sum_{k \in Z^3} f_k \varphi(x - k) \qquad x \in R^3 \tag{5.1.1}$$

where the value $f(x)$ is the linear combination of samples f_k that are evaluated at integer coordinates k, and φ is the reconstruction filter. For computational efficiency the reconstruction filter should have finite support. Such filters will always introduce artifacts into the reconstructed signal that can be classified as post-aliasing, smoothing and ringing (Marschner & Lobb 1994, Mitchell & Netravali 1988).

One of the few user studies that perceptually evaluate the reconstruction filters is that of Mitchell & Netravali (1988). Images displaying typical blurring, ringing and anisotropy artifacts were shown to participants. Test images were then generated using a range of parameters from the BC family of splines and these were compared to the artifact images with the participants being asked to match which of the artifacts the test images most exemplified. Their results determined that different parameters of the same family of spline filters can introduce very different perceptual artifacts.

5.2 Experimental Method

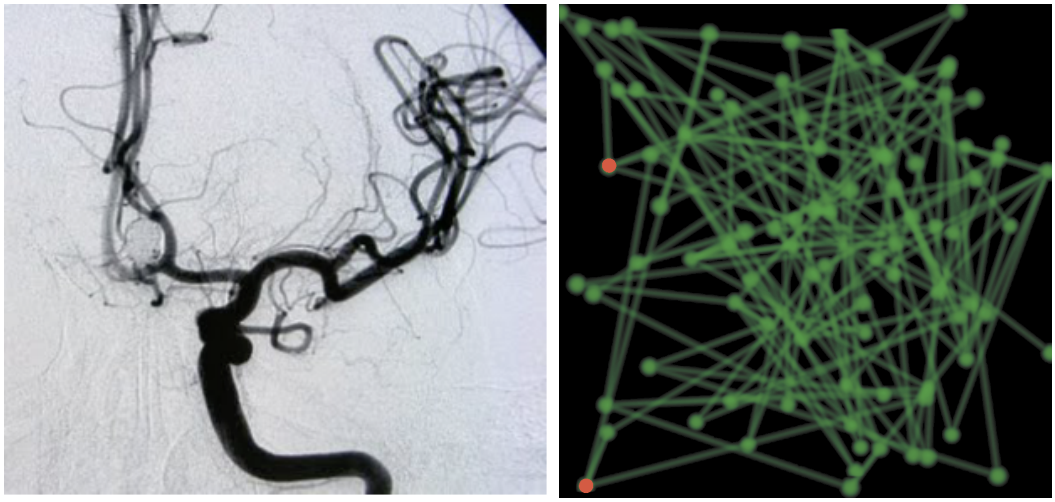
5.2.1 Path-Tracing Setup

In deciding upon a task to perform, there are two key requirements that must be fulfilled. First as DVR is commonly used in the medical domain, the task should be relevant to

the medical profession. Yet, there are no standard viewing procedures that can be used, a problem noted by prior studies (Ropinski et al. 2006). Further, as the participants will be novices, it is required that the task itself does not require medical expertise. Angiography images containing overlapping and connecting vessels are structurally similar to interconnected-graphs, as such a path-tracing task whereby participants must find a path between nodes in a graph can be used. Such tasks have been used in prior experiments due to their similarity with the task of interpreting vascular data sets (Beurden et al. 2012, van Beurden et al. 2011, Ware & Franck 1996).

The setup follows that of previous experiments using graph based-networks (Ware & Franck 1996, van Beurden 2010) with the difference being the use of DVR to produce the images. In each image a graph was displayed with nodes connected via arcs. The participant had to determine whether two nodes highlighted in red were connected via a path of two. For each image there was a 50% chance of either the highlighted nodes being joined by a path of two or there being a path of a different length. If the participant believed there was a path they were instructed to press the 'y' key on the keyboard else to click the 'n' key. The nodes were divided into three equal sized groups, two of which were labeled as leaf nodes and the third labeled as intermediate nodes. In the stereoscopic tests the intermediate nodes were placed on the zero-disparity plane, leaf group 1 were placed within the volume so that they were displayed at a distance of $50mm$ in front of the display and leaf group 2 were placed at a distance of $50mm$ behind the screen. These depths were chosen as they are within a range that allows the image to be viewed with minimal discomfort (Jones et al. 2001). The monoscopic test used the same node layouts with only the perspective cue used. The x, y position of each node within the three layers were randomly distributed. The arcs between each node were determined by connecting each node in both leaf groups to two randomly selected intermediate nodes.

Aside from the changes of the rendering method used between this experiment and prior work (Ware & Franck 1996, van Beurden 2010), there were two more key differences. Firstly, the nodes that act as junction points between the arcs were given the same density value as the arcs. This resulted in them being rendered with the same colour as the arcs connecting the nodes. The reasoning is that in typical blood vessel data sets the junctions where the vessels branch are part of the vessel itself and will therefore have a similar density to the rest of the vessel. The second difference was that with the transfer function (TF) used the nodes and arcs were rendered semi-transparently with the transparency



(a) MRA image (b) Task image

Figure 5.1: Comparison between a rendering of a magnetic resonance angiography (MRA) dataset in Fig (a) (van Beurden et al. 2011) and the task stimulus in Fig (b) with two nodes highlighted red.

dependent upon the density of the object. This produces images that are typical of DVR and reflects prior experiments in the stereoscopic DVR literature (Cho et al. 2012, Kersten 2006, Boucheny et al. 2009). For the purposes of this experiment the number of nodes have been set at 90, a number consistent with previous trials. Figure 5.1 shows an example of the task stimulus compared to a typical angiography blood vessel data set.

5.2.2 Reconstruction Filters to be Analysed

The choice of reconstruction filter used for the experiment is motivated by the requirement of satisfying the interpolation constraint as well as using filters that cover a range of quality. First, the trilinear interpolation scheme has been chosen due to its simplicity and generally acceptable results. Second, the Catmull-Rom spline has been selected as it is the only BC-spline that is interpolating and is used when high-quality rendering is required. Finally the interpolating B-Spline represents one of the highest quality interpolation schemes that can be used. The chosen interpolation schemes are in common usage and exhibit a wide range of quality in terms of the error function notation (N-EF) of Möller et al. (1996) and their smoothness ranging from C^0 to C^2.

Trilinear

Requiring only eight neighboring sample points, trilinear interpolation is the simplest reconstruction scheme capable of recreating a continuous function from a discrete set of 3D samples and is a C^0, 3-EF filter.

Catmull-Rom

The Catmull-Rom spline is categorised under the family of BC-splines with parameters B=0 and C=0.5. Being part of the BC category of splines the filter is a cubic spline, with a support of four in the 1D case. The Catmull-Rom spline is given as the function below (Mitchell & Netravali 1988):

$$Catmull(x) = \frac{1}{6}\begin{cases} (12-3)|x|^3 + (-18+3)|x|^2 + 6 & 0 < |x| < 1 \\ -3|x|^3 + 15|x|^2 + -24|x| + 12 & 1 \leq |x| < 2 \\ 0 & \text{otherwise} \end{cases} \tag{5.2.2}$$

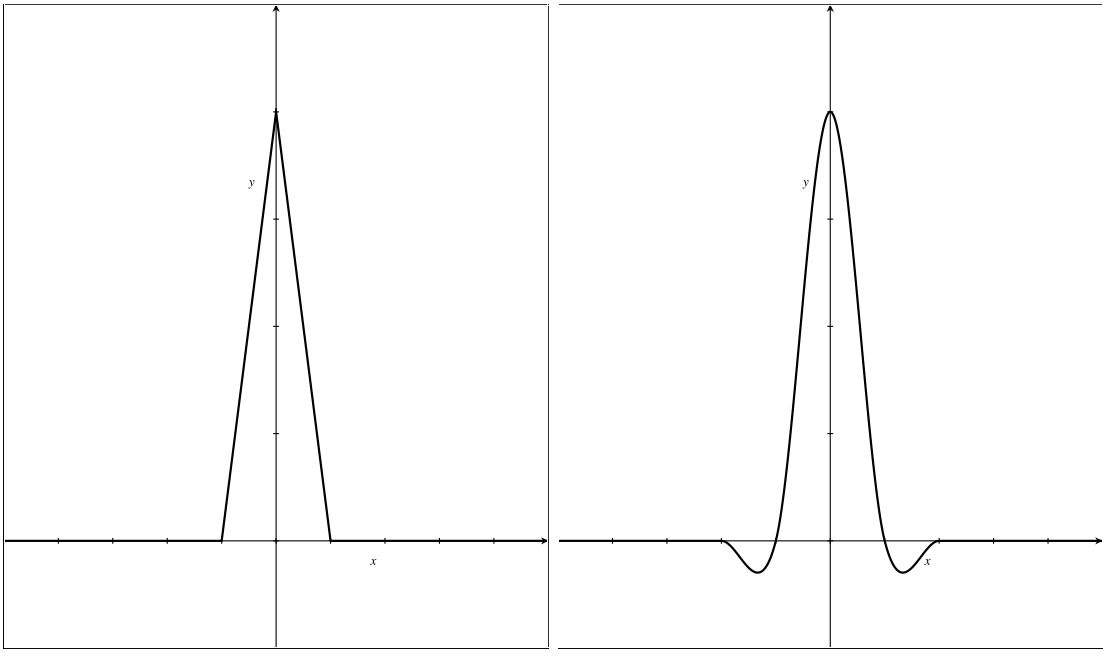
The filter is C^1 continuous resulting in discontinuous curvature when a signal is reconstructed. In terms of accuracy the Catmull-Rom spline is a 3-EF filter. The spatial response of the Catmull-Rom spline can be seen in Figure 5.2.

Interpolating B-spline

The cubic B-spline as defined in Equation 5.2.3 has a history of being used for volume rendering in order to approximate sample points between existing grid points (Marschner & Lobb 1994). The regularity of the cubic B-spline function is C^2, meaning that the function can be differentiated 2 times. Like the Catmull-Rom interpolation scheme, in the 3D form the kernel uses a total of 64 neighbouring points in order to calculate a single sample point. The B-Spline function is given as the function below:

$$B^3(x) = \begin{cases} \frac{2}{3} - \frac{1}{2}|x|^2(2-|x|) & 0 \leq |x| < 1 \\ \frac{1}{6}(2-|x|)^3 & 1 \leq |x| < 2 \\ 0 & 2 \leq |x| \end{cases} \tag{5.2.3}$$

The spatial responses of the B^3 kernel can be seen in Figure 5.3. Such functions are used in signals that are inherently noisy due to its natural smoothing behaviour and has the property that all values at the centre of the aliasing spectra are equal to zero, eliminating the reconstruction error of sample frequency ripple (Csébfalvi 2008). Although the kernel



(a) Linear (b) Catmull-Rom

Figure 5.2: The 1D linear interpolation scheme compared to the Catmull-Rom interpolation scheme. Note the non-linearity of the Catmull-Rom spline and the negative lobes, similar to the sinc function.

has been used directly in volume rendering for interpolating values between data points, there is a significant problem in doing so. The B^3 and higher degree B-splines are not interpolating and therefore do not satisfy the interpolation constraint meaning that any reconstructed volumes will not take the original data values at sample positions lying on integer positions of the volume. In visualizations this produces a significant blurring of the images.

Via a process known as *generalized interpolation* (Blu et al. 1999) the interpolating B^3-spline can be derived from the approximating B^3-spline. Equation 5.1.1 is in this case modified to produce:

$$f(x) = \sum_{k \in Z^3} c_k \varphi(x - k) \qquad x \in R^3 \tag{5.2.4}$$

where the coefficients c_k are computed from the sample data f_k after applying the inverse of the filter. The interpolating cubic B^3-spline produces a C^2 continuous 4-EF filter having a continuity and accuracy greater than the Catmull-Rom spline.

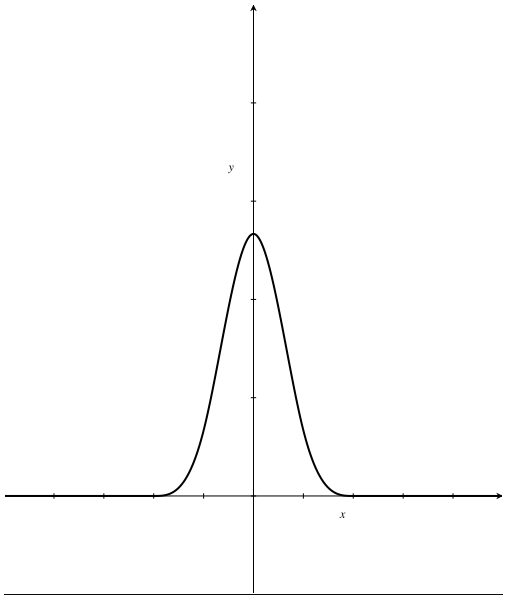


Figure 5.3: The B-Spline approximation kernel.

5.2.3 Stimulus Generation

As per the previous chapter a voxelisation method was required to generate the simulated data sets with the aim of producing the minimum of artifacts during this stage. In the previous chapter a voxelisation method based on the work of (Sramek & Kaufman 1999) was used. In order to further reduce the chance of artifacts that can occur during this stage an updated voxelisation method has been used to generate the graph data sets. For this experiment the *vxtRL* voxelisation software (Novotny et al. 2010) that provides an implementation of the truncated distance fields method of voxelisation has been used. For all graphs shown the volume resolution has been set at $(256 \times 256 \times 256)$.

An in-house software-based raycaster was used to generate the images with the sample rate set above that of the Nyquist rate of the data set. The same transfer function was used to generate all images with a constant green value of 0.35 and an increasing linear opacity of $0 - 0.6$.

5.2.4 Participants

18 participants were recruited to take part in the experiment. In total there were 4 women and 14 men with ages ranging from 18-27 and a mean age of 21, each of the participants were paid £5 to take part in the experiment. All participants taking part in the study were screened for vision by using the Bailey and Lovie Snellen chart and for stereo-vision

using the Titmus fly stereo test. Some of the participants did have prior experience with 3D displays, however all of the participants were novices in regard to the nature of the task and none had any medical training or experience with studying medical data sets.

5.2.5 Equipment and Viewing Conditions

A True3Di 24″ HD-SDI Monitor was used for displaying the graph images in both the stereoscopic and monoscopic test setups. The display provides a WUXGA resolution of 1920×1200 pixels for each eye with a refresh rate of $60Hz$. In order to view the 3D images linear polarized glasses were required to be worn. The glasses were worn in both the stereoscopic and monoscopic viewing conditions so as to remove a possible source of bias.

Participants were placed at a distance of $60cm$ from the display. The experiment took place in a darkened room with the light levels kept consistent throughout the experiments.

5.2.6 Procedure

Upon taking part in the experiment each participant was required to read written instructions as well as provide a signature marking that they understood the instructions and were happy to proceed with the experiment. Each participant were then screened using the Lovie Snellen chart and for stereo-vision using the Titmus fly stereo test. It was made clear to the participants that although accuracy and response times were recorded for each image, accuracy should be given precedence over response time.

A trial test was performed with randomly generated simplified graphs containing only 30 nodes generated with the 3 reconstruction filters. The graphs were displayed randomly either stereoscopically or monoscopically. During the trial test when the participants made a choice a screen was presented to them stating whether the choice was correct. The trial period continued until the participant felt comfortable enough to proceed with the full trial. Accuracy and response times were not recored during the trial.

After the trial period the full experiment began. There were three trials for the stereoscopic and monoscopic presentations with the three trials corresponding to the three reconstruction filters for a total of 6 trials. For each of the 6 trials 20 images were presented, after each trial a short break was provided in order to reduce the impact of fatigue. The exact ordering of the trials depended upon which group the participant was assigned to according to the ordering of the Latin Square design used. Although the task duration was

somewhat dependent upon the individual participant, in general the experiment lasted for approximately 60 minutes.

5.3 Hypothesis

Our goal is to understand the influence of reconstruction when measuring accuracy and response time for a path-searching task. Our first question is therefore: What are the benefits to using higher quality reconstruction methods when performing spatially complex tasks with vascular-like data sets? Secondly we are interested in how task performance may differ between display types with the different reconstruction methods. Our second question is then: Does the display type have an impact on results when different reconstruction methods are considered?

In the literature, the assessment of reconstruction filters for volume rendering has taken place either via juxtaposition with images produced by different filters, typically using the Marschner and Lobb data set or via some metric. Aliasing can lead to inaccuracies when stereo depth perception is required (Pfautz 2000). Further when excessive blur is introduced into a stereo image the ability to perceive depth correctly diminishes as the amount of blur increases. It may be expected then that reconstruction filters with poorer pass-band behaviour and therefore greater smoothing will exhibit a decrease in accuracy and an increase in response time during the experiment. Reconstruction filters that more accurately reflect the sinc filter are expected to produce an increase in task performance.

5.4 Results

The response latencies and accuracy rates were subjected to two-way repeated measures of analysis of variances (ANOVAs). The reconstruction method and display setup, either monoscopic or stereoscopic, were the independent variables and the accuracy and response time were the dependent variables. The mean correct responses for each condition are shown in Table 5.1 and the mean response latencies are shown in Table 5.2 these are represented visually in Figure 5.4. A summary of the ANOVA results can be seen in Table 5.3.

Table 5.1: Mean Accuracy (percentage)

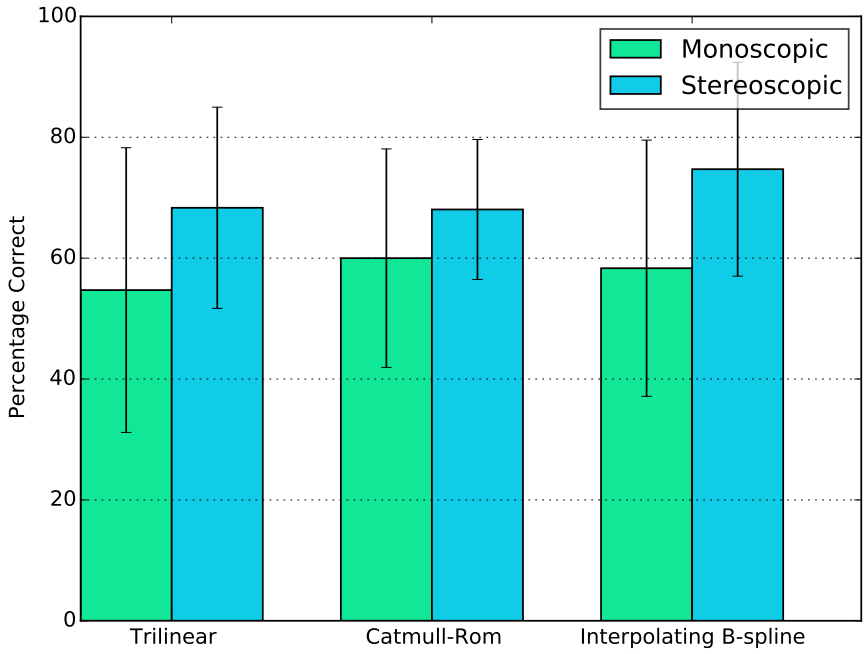
	Monoscopic		Stereoscopic	
Reconstruction	Mean	SD	Mean	SD
Trilinear	54.720	23.556	68.335	16.646
Catmull-Rom	60.000	18.075	68.055	11.586
Interpolating B-spline	58.335	21.197	74.720	17.699

Table 5.2: Mean Response Latencies (seconds)

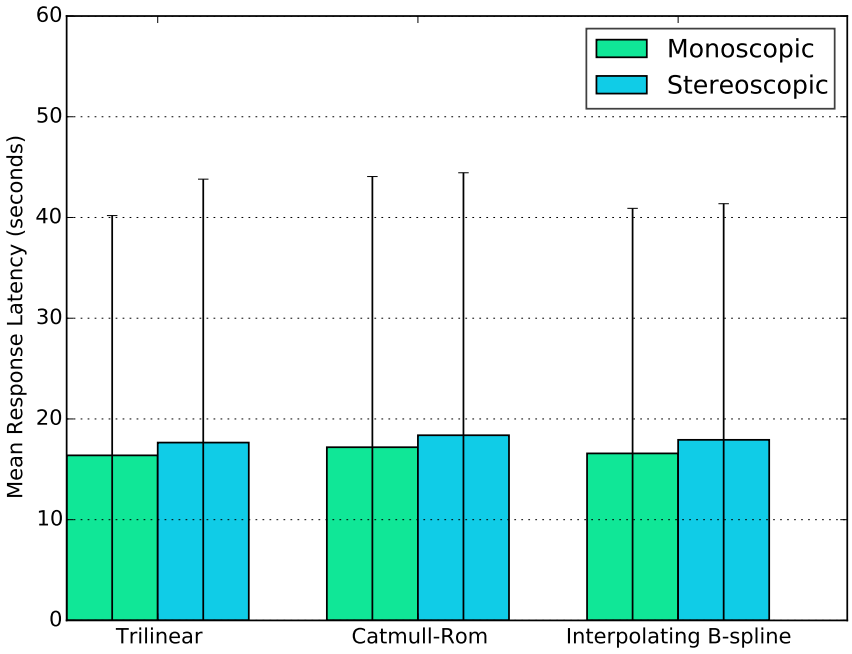
	Monoscopic		Stereoscopic	
Reconstruction	Mean	SD	Mean	SD
Trilinear	16.385	23.804	17.654	26.152
Catmull-Rom	17.193	26.871	18.382	26.057
Interpolating B-spline	16.579	24.329	17.926	23.441

Table 5.3: Summary of accuracy results. Asterisk denotes significant p-values.

	p-value	
Condition	Accuracy	Response Time
Display	0.001*	0.067
Reconstruction	0.030*	0.485
Display x Reconstruction	0.367	0.993



(a) Accuracy



(b) Respnse Time

Figure 5.4: Performance for path-searching experiment with standard deviations.

5.4.1 Display

For accuracy a two-way analysis of variance collapsed across the interpolation method determined that there was a significant effect from monoscopic to stereoscopic display setup with $F(1, 17) = 22.586$, $p < 0.001$. There was an increase in accuracy from 11.648 ± 0.430 to 14.074 ± 0.372 a statistically significant increase of 2.426 (98% CI, 1.349 to 3.503).

A repeated measure two-analysis of variance collapsed across the interpolation method was used to determine whether there were any statistical significance between response times of display types. There was a weakly significant effect from monoscopic to stereoscopic display setup with $F(1, 359) = 3.366$, $p = 0.067$. The response time increased from monoscopic to stereoscopic with the averages increasing from 16.719 ± 1.214 to 17.987 ± 1.203 a marginally significant increase of 1.269.

5.4.2 Reconstruction Filter

A two-way analysis of variance collapsed across the display type was used to determine if there were statistically significant differences in accuracy between the interpolation schemes used. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated $\chi^2(2) = 0.098$, $p = 0.952$. The interpolation scheme used elicited statistically significant changes in accuracy with $F(2, 34) = 6.800$, $p = 0.030$ with mean accuracy increasing from 12.306 ± 0.343 for trilinear interpolation to 12.806 ± 0.339 for Catmull-Rom interpolation and to 13.472 ± 0.397 for interpolation B-spline. Post hoc analysis with a Bonferroni adjustment revealed that accuracy significantly increased between trilinear and interpolation B-spline reconstruction with $p = 0.004$ an increase of 1.167 (95% CI, 0.356 to 1.977). There was no significant differences between the accuracies for trilinear interpolation and Catmull-Rom or interpolating B-spline and Catmull-Rom. A further one-way ANOVA considering only the stereoscopic trials determined that there was a weakly significant effect between interpolation types with $F(2, 34) = 2.850$, $p = 0.072$. Bonferroni pair-wise comparison found there were no significant differences between individual pairs of interpolation methods.

For the assessment of interpolation on response time the Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated with $\chi^2(22) = 2.462$, $p = 0.485$. The ANOVA determined that there was no statistical significance between interpolation type and response time with $F(2, 7128) = 0.724$, $p = 0.485$.

5.4.3 Display with Reconstruction Filter

A two-way repeated measures ANOVA was used to examine the effect of display type and interpolation scheme on accuracy. Mauchly's Test of Sphericity had been violated with $\chi^2(2) = 10.489$, $p = 0.005$, therefore a Greenhouse-Geisser correction was applied ($\epsilon = 0.675$). No statistically significant changes were found with $F(1.351, 22.960) = 0.953$, $p = 0.367$.

A two-way repeated measures ANOVA was conducted that examined the effects of display type and interpolation scheme on response latency. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated with $\chi^2(2) = 3.759$, $p = 0.153$. No statistical significance was found with $F(2, 718) = 0.007$, $p = 0.993$.

5.5 Discussion

Reviewing the hypothesis the results confirmed that higher quality reconstruction methods can produce higher accuracy results for spatial tasks, further stereoscopic displays can produce improved results for these tasks.

In comparing the benefit of stereoscopic displays, the significant increase of accuracy of 12.1% is comparable with 11.4% increase found in a prior experiment to determine the required number of stereo view points (Hassaine et al. 2010). However the accuracy rates are below those of prior studies with similar graph complexities in both display conditions (Ware & Franck 1996, Hassaine et al. 2010, Ware & Mitchell 2005). The results are comparable to those found in Cho et al. (2012) whereby participants were required to determine the correct ordering of a set of semi-transparent volume rendered cylinders. This implies an increased level of difficulty when transparency is introduced in graph networks.

When comparing reconstruction filters the results suggest that there was a significant impact on accuracy when display type is not considered, specifically between trilinear reconstruction and interpolating B-spline. Further investigation comparing the stereoscopic trials implies that there may be differences between the reconstruction filters, although the exact nature is not clear. The display and reconstruction filter results yielded no significant impact on accuracies. An explanation might be, as mentioned above the complexity of the task that may be obscuring further differences causing a flooring effect. The low accuracy percentages when compared to prior experiments provides evidence to support

this. Although the trends suggest that reconstruction filters of higher quality, according to the N-EF notation, may produce more accurate results further research is required to determine the exact nature of the effect with a range of task complexities and a broader range of reconstruction filters.

Regarding response latency, only the display type had a marginally significant effect with the stereoscopic trials producing larger response latencies. Although these results are different to previous path-searching studies that suggest that stereoscopic displays decrease response times (Ware & Franck 1996, Hassaine et al. 2010) it is in line with studies from the stereoscopic volume rendering literature (Ropinski et al. 2006, Laha et al. 2012). A possible explanation is that having multiple layers in depth may extend search time because each depth must be searched separately.

5.6 Conclusions

In this chapter we have examined the effect of reconstruction filters in DVR with stereoscopic and monoscopic display conditions for a path-searching task. Results showed that stereoscopic display of the data significantly improved accuracy with no significant impact on response time. The results suggest that the reconstruction filter can have an effect on tasks that require spatial understanding with the general trend implying that as the reconstruction filter quality increases accuracy also increases. In relating the practical implications of this study to the medical domain it is apparent that the filter for rendering vascular-like data sets cannot be made arbitrarily and should depend upon the context within which the rendering is to be used. Although further research is required to clarify the results with a wider range of filters and tests, the results suggest a possible explanation for the wide range of results found in experiments involving stereoscopic volume rendering.

Although an effect was found on the choice of reconstruction filter, the effect was subtle. In order to evaluate these further, it is necessary to test a more varied set of filters that have different ranges of post-aliasing and smoothing. Thus, in the following chapter we evaluate an extended range of filters. From discussions with participants, it was indicated that some bias may exist in the results. From analysing the responses, this was confirmed with 62.344% of the responses being ‘No’. This bias is intriguing, and in the next chapter we use techniques from Signal Detection Theory to assist in analysing the bias and sensitivity of responses.

The study presented does have limitations, the primary of which is the inability to

separate the image artifacts of post-aliasing and smoothing on the response of the task. Smoothing is known to effect the results of stereoacuity tests and aliasing can also have an impact on tasks requiring depth perception (Pfautz 2000, Costa et al. 2010). Further work may seek to classify the reconstruction filters using the frequency domain metrics of Marschner & Lobb (1994) to provide additional insight. A second limitation is the use of simulated data sets that express a regularity in terms of vessel thickness, junction connections and lack of noise that are not found in typical volume data sets.

Chapter 6

Reevaluating Reconstruction Filters for a Path-Searching Task in 3D

This chapter is based on work in 'Reevaluating Reconstruction Filters for a Path-Searching Task in 3D' (Roberts & Ivrissimtzis 2016*b*). It is an experimental study evaluating a range of reconstruction filters, two transfer functions and two display conditions on a task requiring spatial understanding.

6.1 Introduction

In this chapter we extend the evaluation of reconstruction filters from Chapter 5 for a path-searching task using DVR. In Chapter 5, we evaluated the results of a path-searching task with DVR on three reconstruction filters. The results suggested that reconstruction filters can have a significant impact on rates of accuracy for this type of task specifically between the trilinear and interpolating B-spline reconstruction filters.

We compare five DVR reconstruction filters and two opacity parameters in terms of accuracy and response times when displayed stereoscopically and in 2D for a path-searching task. The scene consists of a graph with spherical nodes and cylindrical edges, with a certain amount of added Perlin noise. The task is to specify whether two nodes, highlighted with the colour red, are connected with a path of length two or not. We seek to answer the following questions: 1) How does the opacity, controlled by the choice of transfer function, relate to task performance? 2) How does post-aliasing and smoothing,

controlled by the choice of reconstruction filter, relate to task performance? 3) How do the results between stereoscopic and monoscopic displays compare?

The analysis of the results showed that our experiment was eliciting biased responses from the participants. More specifically, a No answer was more likely than a Yes answer with 63.14% of the responses in 3D and 56.58% of the responses in 2D being No answers. Such bias, which is common in Yes-No experiments, is not necessarily a negative aspect. For example, a surgeon wanting to reduce unnecessary operations may have a preference towards saying no, whereas when detecting a tumor may say yes more often. However, as a result of the bias in our experiment, we found that the standard statistical analysis of the overall accuracy rates is inadequate and techniques from *signal detection theory*, give insights that other techniques miss.

Our analysis shows that in 3D displays the level of opacity has the greatest effect on task performance, with the accuracy increasing with opacity. The overall effect of the choice of reconstruction filter on accuracy rates was not strong, however, the choice of filters with large amounts of post-aliasing had a strong effect in maximising the number of correct rejections. Using signal detection theory techniques, we also found that a choice of in-between filters, which avoid large amounts of both post-aliasing and smoothing, had a strong effect in maximising the ability of the participants to discriminate between connected and not-connected nodes. When the whole experiment was duplicated in 2D we found that task performance was significantly lower compared to 3D.

A first implication of the results is that if the maximisation of correct rejections is the primary concern, e.g. during a process of selection by elimination, then filters with higher post-aliasing should be preferred. On the other hand, if the ability to discriminate between connected and not-connected nodes is the primary concern, then filters with a better balance between post-aliasing and smoothing should be preferred. We note that this second observation stems from the signal detection theory analysis and possibly due to the bias in the data, does not translate into a strong effect of the reconstruction filter on the overall accuracy rates of the experiment.

A second implication of the results of the chapter is that, unless experimental designs eliciting unbiased responses are employed, such as m-Alternative Forced Choice (mAFC) experiments, signal detection theory should be applied to measure the bias and then explicitly take it into account in the analysis. Finally, if for any reason signal detection theory cannot be the analytical tool of choice, the experimental results should nevertheless be

presented in a way that bias can be computed. In the case of binary decision experiments in particular, hits and correct rejections should be separately reported, rather than just accuracy rates.

6.1.1 Evaluations of Stereoscopic Volume Rendering

A number of studies have concluded that stereoscopy can assist in certain tasks with DVR that require spatial understanding, however the results are not unanimous or consistent. A discussion of the studies can be found in Chapter 5, Section 5.1.1.

6.1.2 Signal Detection Theory

The path-searching experiment reported in this study can be classified as a Yes-No experiment. Historically the analysis of such experiments have been limited to the statistical analysis of raw accuracy results and in some cases response times of each participant. By applying techniques from Signal Detection Theory (Macmillan & Douglas Creelman 2004), a wider range of analysis can be performed, providing a different insight into the results gathered. In particular, different types of errors can be evaluated, as well as the sensitivity and bias of the participants, providing a more finer analysis of the results than statistical analysis of raw accuracy alone.

Signal Detection Theory originated as a method of analysing experiments where auditory or visual stimuli have to be distinguished from some background noise. Since this initial use, it has been applied to the analysis of memory tasks, cognition as well as tasks within the medical domain (Abhari et al. 2015).

6.2 Experimental Methodology

6.2.1 Path-Searching Task Description

In each image presented to the participants a direct volume rendered graph is displayed. The task of the participant is to determine whether two highlighted nodes in the graph are connected by a path of two. Each trial is set so that the chance of the nodes being connected by a path of two is 50%. The ‘y’ key on the keyboard is used for the participant to indicate there is a path of two where as the ‘n’ key is used to indicate they cannot see a path.

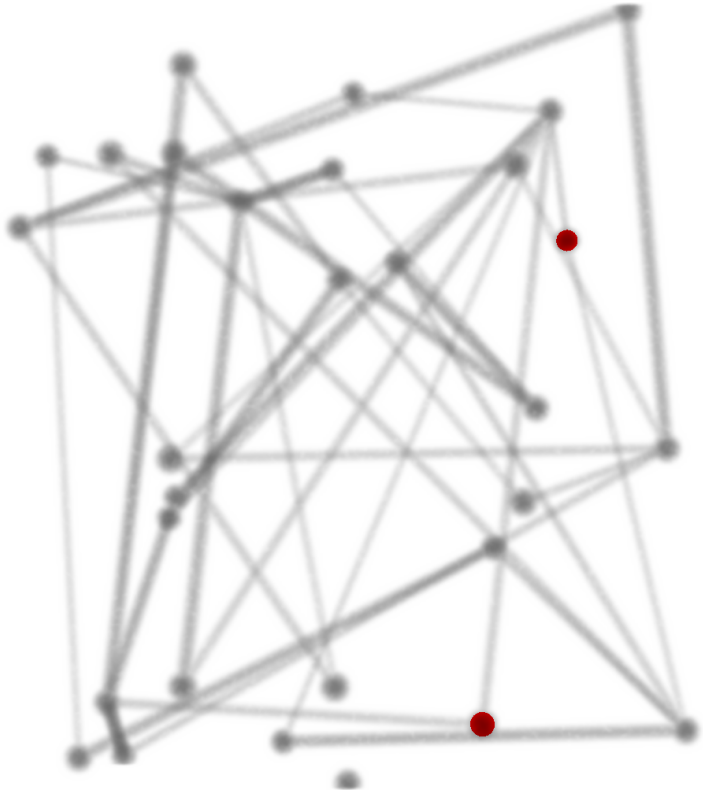


Figure 6.1: Example of a test graph image displayed to participants. Thirty nodes are displayed with two highlighted nodes that are connected by a path of two.

The structural layout of the graph follows that of Ware & Franck (1996), Hassaine et al. (2010). The number of nodes in all trials was set to 90 and they were evenly divided into three sets, two leaf sets and a single set of intermediate nodes. When viewed stereoscopically, the intermediate nodes were displayed on the zero-disparity plane without depth, the first leaf group was displayed at $50mm$ in front of the display and the second leaf group at $50mm$ behind the display. In the 2D experiment, all nodes were displayed without depth on the screen plane. The x and y coordinates were randomly distributed.

The chosen path-searching task is similar to the search for aneurisms in real-life angiography data sets. Both of the tasks require a high-degree of spatial understanding and require the observers to understand the layouts of the scene (Grosset & Schott 2013, Ropinski et al. 2006). The layout and structure of angiography images and those used in the path-searching task are not commonly found in every-day situations and so participants are unable to rely on prior experience to immediately comprehend the layout of the images (Ropinski et al. 2006). For illustration, an example test image with only 30 nodes is shown in Figure 6.1.

6.2.2 Stimulus Generation

The graph data sets were generated using the voxelisation package *vxtRL* (Novotny et al. 2010) with resolution $256 \times 256 \times 256$. In order to produce images that are representative of blood vessel data sets, two features have been used. The first is that the lines connecting the nodes in the graph have been randomly assigned a diameter from the choice of [0.0015, 0.003, 0.006] where the size of the scene in normalized coordinates is [0, 1]. These diameters represent fine, medium and thick blood vessels.

Secondly, volumes generated via medical scanners and other devices often exhibit some form of noise, for example in volumes created using ultrasound scanners this is a primary cause of concern. In order to simulate this property the generated graph volumes are modulated by a texture with the resolution of $(256 \times 256 \times 256)$. Each point in the texture is given a value according to the Perlin noise function where the parameters a and b that control the frequency and harmonics of the noise have been chosen as $a = b = 2$. By modulating the generated graphs with the noise the produced volumes have an irregularity that approximates those of natural data sets.

6.2.3 Reconstruction Filters to be Evaluated

The five reconstruction filters to be evaluated were chosen for their wide range in the amount of smoothing and post-aliasing they introduce, as well as being in common usage in DVR. In order to quantitatively assess the amount of smoothing and post-aliasing each filter exhibits, the frequency domain metrics of Marschner and Lobb have been computed (Marschner & Lobb 1994).

There, smoothing is assessed by measuring the energy the filter extrudes in the pass-band:

$$S(h) = 1 - \frac{1}{|R_N|} \int_{R_N} |\hat{h}|^2 dV \tag{6.2.1}$$

where h is the filter to be measured, $\hat{h}$ is the frequency domain representation of h, R_N is the Nyquist region and $|R_N|$ is the volume of R_N. Post-aliasing is assessed by measuring the energy the filter extrudes in the stop-band:

$$P(h) = \frac{1}{|R_N|} \int_{\overline{R_N}} |\hat{h}|^2 dV \tag{6.2.2}$$

where $\overline{R_N}$ is the complement of R_N.

The *trilinear* interpolation scheme is the simplest and most common filter capable of recreating a continuous function. It produces images of acceptable quality and as it is

implemented in hardware in GPUs it has become the *defacto* standard reconstruction filter for DVR. The *Catmull-Rom spline* satisfies the interpolation constraint whilst balancing smoothing and aliasing. The *cubic B-spline* is an approximating filter exhibiting minimal post-aliasing at the cost of significant smoothing. Due to this it is used primarily in DVR when there is a large amount of noise present. Derived from the approximating B-spline filter via the use of generalized interpolation (Blu et al. 1999), the *interpolating B-spline* has a higher degree of smoothing and greater accuracy than the trilinear and Catmull-Rom filters whilst satisfying the interpolation constraint.

The choice of the last filter is motivated by the desire to test a reconstruction filter that exhibits high post-aliasing with minimal smoothing. A number of choices are available that include the pass-band optimal filters of Carlbom (1993) as well as the premultiplied cubic filters of Csébfalvi (2008). Such filters are however not in common usage in DVR. Instead, much research has gone into the study of windowed-sinc filters whereby a finite window with some radius is applied to the ideal *sinc* reconstruction filter to produce a more practical filter. By choosing an appropriate window function and radius a considerable degree of control over the frequency domain artefacts can be exercised. The Welch window function is defined as:

$$Welch_{\tau}(x) = \begin{cases} 1 - (\frac{|x|}{\tau}) & |x| \leq \tau \\ 0 & \text{otherwise} \end{cases} \tag{6.2.3}$$

where τ is the radius of the window and for the purposes of this experiment is set to 4. This produces a reconstruction filter that has low smoothing, less than that of the interpolating B-spline, whilst exhibiting significant post-aliasing. The smoothing and post-aliasing properties of each reconstruction filter according to the Marschner and Lobb metrics are given in Table 6.1 and displayed in Figure 6.2. Zoomed in images of the graph stimulus used for the experiment generated with the different reconstruction filters are shown in Figure 6.3 and the differences between two reconstruction filters are highlighted in Figure 6.4 for an opaque rendering of the graph data set.

For each reconstruction filter in the experiment, two opacity levels, implemented via transfer functions, have been used to produce two test images that are typical of DVR. The first produces semi-transparent renderings with a white colour value of 0.35 and an increasing linear opacity of $0 - 0.6$. The second transfer function has the same colour settings and a step function with 0.95 at non-zero values for opacity, producing more opaque images. The choice of transfer functions is motivated by two common styles of

Table 6.1: Marschner and Lobb Metrics

Reconstruction Filter	Post-aliasing	Smoothing
Trilinear	0.065	0.764
Catmull-Rom	0.107	0.567
Interpolating B-Spline	0.055	0.409
B-Spline	0.016	0.906
Welch Windowed Sinc	0.202	0.392

volume rendering. Semi-transparent rendering is typically used in DVR to allow multiple layers to be visible, while in iso-surface rendering the surfaces of interest are made opaque. To be able to compare the effects of the reconstruction filter and opacity on different display methods the experiment has been performed twice once in 3D and once when presented in 2D without depth.

6.2.4 Equipment and Viewing Conditions

As in the previous chapter, a True3DI 24" HD-SDI monitor with a resolution of 1920×1200 pixels for each eye and a refresh rate of $60Hz$ was used to display the images for the experiment. Participants were required to wear linear polarized glasses during both the 3D and 2D experiments and were placed at a distance of $60cm$ from the display. Light levels were kept consistently low throughout the experiments.

6.2.5 Participants

In total 60 participants were recruited to take part with 30 performing the experiment in 3D and 30 in 2D. There were 31 women and 29 men with ages ranging from $18-39$ and a mean age of 20.75. Each participant was paid £7.50 to take part in the experiment. All participants were screened for vision using the Bailey and Snellen chart and for stereo-vision using the Titmus fly stereo test. All participants were informed that accuracy and response time would be recorded during the task. Whilst some of the participants had experience with stereoscopic displays, they were unfamiliar with the DVR image generation process and had minimal knowledge of reconstruction filters.

The use of novices follows that of similar evaluations of volume data sets (Laha et al.

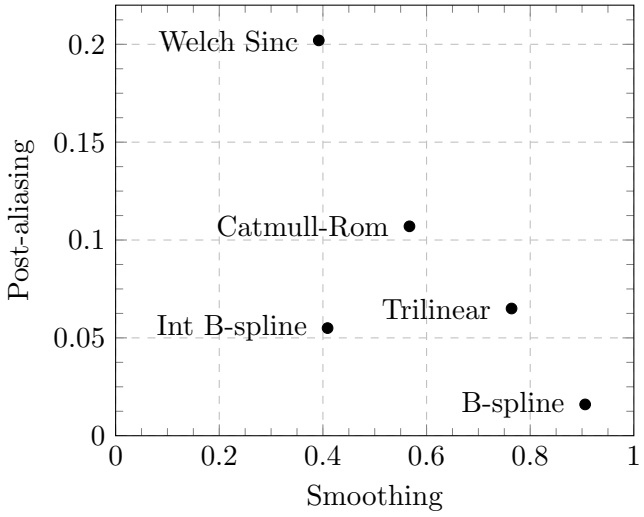


Figure 6.2: Marschner and Lobb smoothing and post-aliasing metrics for the reconstruction filters used in the experiment. Higher values indicate an increased presence of smoothing/post-aliasing.

2012, 2014). Apart from the difficulty in obtaining a significant number of domain experts in volume data, their prior knowledge can obfuscate the results. On the other hand, search tasks have been identified as being relevant to domain experts whilst not requiring specialised knowledge (Laha et al. 2014).

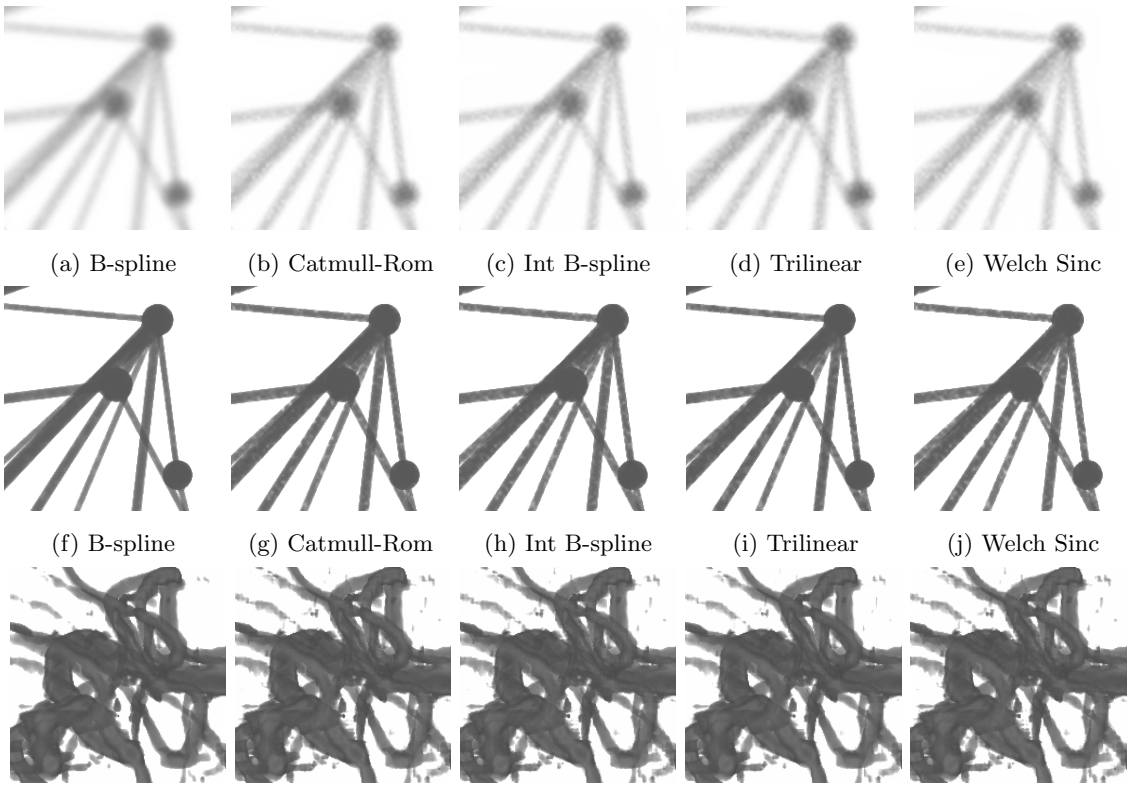
6.2.6 Procedure

Prior to taking part in the full experiment, participants were required to perform a trial test using graphs with a reduced difficulty level containing only 30 nodes for training purposes. The trial test contained six images with three images containing a path of two and three not. In both the 3D and 2D experiments there were five trials for each reconstruction filter per the two opacity levels leading to a total of ten trials. Each trial contained twelve images, after which there was a short break to reduce the risk of fatigue. A Latin Square design was used to determine the exact ordering of the trials.

6.2.7 Hypothesis

Based on the findings of the prior work, our initial hypotheses were:

- Reconstruction filters that exhibit extremes of either blurring or aliasing will display lower overall accuracy rates.

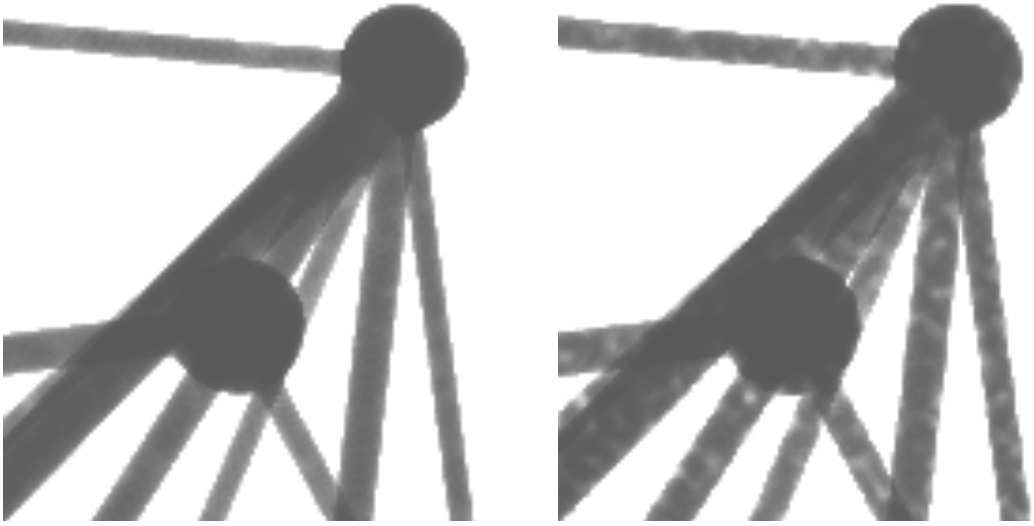


(a) B-spline (b) Catmull-Rom (c) Int B-spline (d) Trilinear (e) Welch Sinc

(f) B-spline (g) Catmull-Rom (h) Int B-spline (i) Trilinear (j) Welch Sinc

(k) B-spline (l) Catmull-Rom (m) Int B-spline (n) Trilinear (o) Welch Sinc

Figure 6.3: Zoomed in view of the graph stimulus and blood vessel data set with different reconstruction schemes. Figs a-e have been generated with the semi-transparent function, Figs f-j have been generated with the opaque transfer function. Figs k-o have been generated with an arbitrary transfer function. Note that the differences between the renderings are subtle.



(a) B-spline (b) Welch

Figure 6.4: Enlarged and zoomed in view of the graph stimulus to highlight the differences between the B-spline and the Welch Sinc reconstruction filter.

- The accuracy rates will be higher in 3D than in 2D.

6.3 Results

Statistical analysis of the experiment are presented in this section. Section 6.3.1 presents the results of the path-searching task with five reconstruction techniques and two opacity levels. Further analysis of the results when categorised into hits, misses, correct-rejections and false-alarms are reported in Section 6.3.2. Section 6.3.3 continues this analysis after applying techniques from Signal Detection Theory to calculate the sensitivity measure per participant as well as participant bias for each condition. The results of matching the task performance to the frequency domain properties of the reconstruction filters are reported in Section 6.3.4. A summary of the results can be seen in Table 6.2.

6.3.1 Overall Accuracy and Response Time

First, the accuracy rates and response times of the 2D and 3D experiments are analysed separately, applying a two-way repeated measures of analysis of variance (ANOVAs) for the within subjects variables of opacity level and reconstruction filter. Next, the combined results of the experiments, are analysed with a one-way ANOVA and any significant p-values are discussed at the end of the relevant subsections. Mauchly's Test of Sphericity

Table 6.2: Summary of results. TF and Rec denote transfer function and reconstruction filter, respectively. Asterisks denote significant p-values, $p < 0.05$.

	3D		2D	
Condition	Accuracy	Response Time	Accuracy	Response Time
		Overall		
TF	0.010*	0.369	0.038*	0.064
Rec	0.576	0.451	0.893	0.915
TF x Rec	0.576	0.451	0.017*	0.758
		Hits (True-Positive)		
TF	0.183	0.623	0.681	0.160
Rec	0.906	0.180	0.928	0.543
TF x Rec	0.451	0.215	0.052	0.476
		Misses (False-Negative)		
TF	-	0.450	-	0.975
Rec	-	0.046*	-	0.018*
TFx Rec	-	0.841	-	0.260
		Correct Rejections (True-Negative)		
TF	0.002*	0.646	0.020*	0.132
Rec	0.621	0.485	0.234	0.894
TF x Rec	0.022*	0.337	0.079	0.904
		False-Alarms (False-positive)		
TF	-	0.112	-	0.352
Rec	-	0.457	-	0.100
TF x Rec	-	0.766	-	0.229

was used to test for sphericity with the Greenhouse-Geisser correction being applied when required.

In the two-way ANOVA, the reconstruction filter and opacity level, classified as either semi-transparent or opaque, were the independent variables and the accuracy and response times were dependent variables. The p-values are summarised in Table 6.2. For each condition, means and standard errors for the correct responses are shown in Figure 6.5a and for the response times in Figure 6.5b.

In both 2D and 3D, the opacity parameter had an effect on accuracy. In 3D, statistical significance was found with $F(1, 29) = 7.657$, $p = 0.010$, partial $\eta^2 = 0.209$. The accuracy increased from semi-transparent (8.033 ± 0.191) to opaque rendering (8.573 ± 0.177). In 2D statistical significance was found with $F(1, 29) = 4.741$, $p = 0.038$ partial $\eta^2 = 0.141$. Again, accuracy increased from semi-transparent (6.787 ± 0.155) to opaque (7.187 ± 0.137). There were no significant effects on response time.

The one-way ANOVA of the display variable showed that participants performed better in 3D than in 2D with $F(1, 29) = 43.288$, $p < 0.001$, partial $\eta^2 = 0.599$ increasing from 2D (6.987 ± 0.114) to 3D (8.303 ± 0.156).

6.3.2 Hits and Correct Rejections

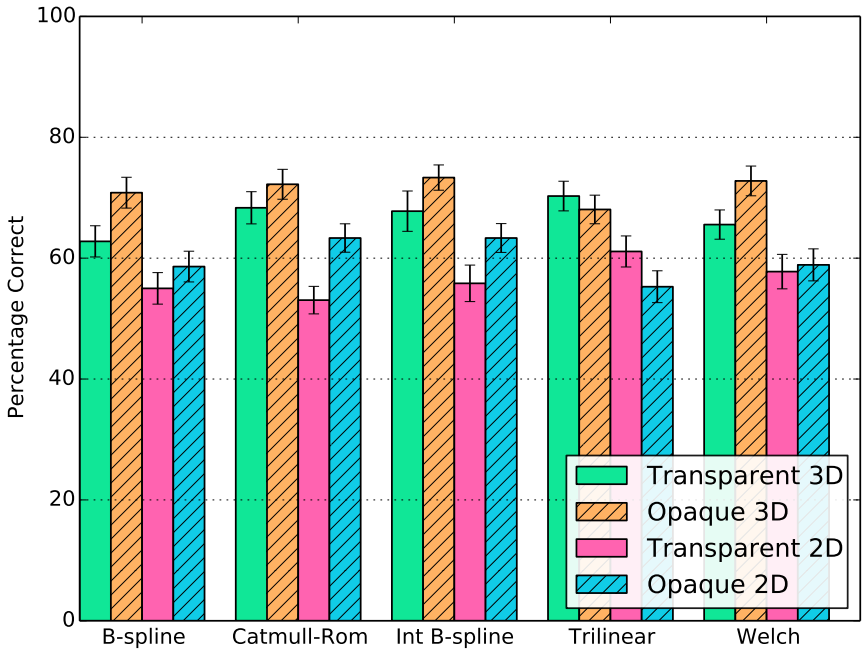
The results of the experiment can be classified into four categories. Using signal detection theory terminology, the four categories are 'Hits', 'Misses', 'Correct Rejections' and 'False-Alarms'. These correspond to True-Positives, False-Negatives, True-Negatives and False-Positives respectively.

The means and standard errors of the accuracy rates can be seen in Figure 6.6. Only the Hits and Correct Rejections are given as the Miss and False-Alarm rates can be calculated as (50%-Hits) and (50%-Correct Rejections).

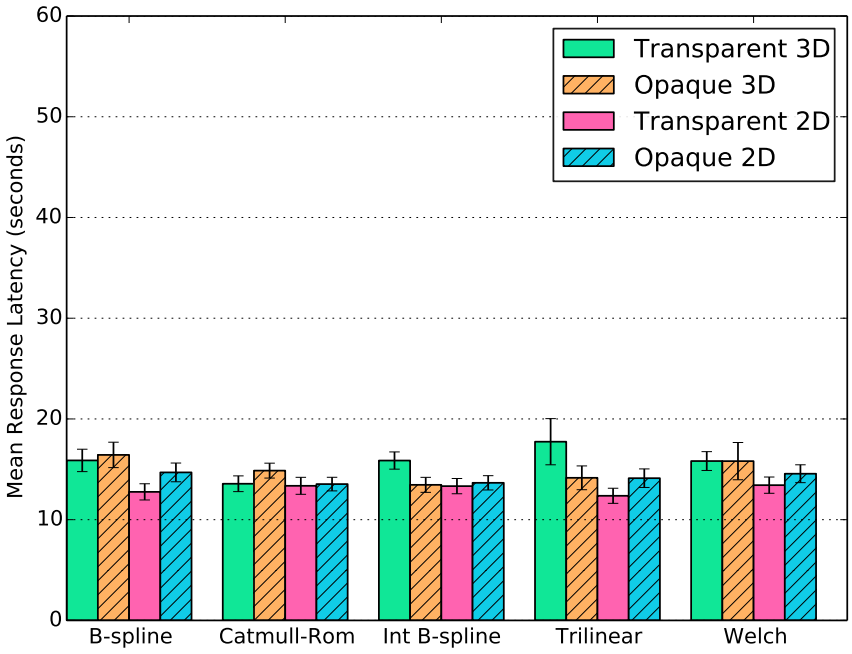
Categorised Accuracy

No statistical significance was found for either the opacity parameter or reconstruction filter on the number of hits in the 2D or 3D experiments.

Regarding correct rejections, there was no effect in the 2D experiment, however, in 3D the opacity parameter was found to have a significant effect. The results were $F(1, 29) = 11.581$, $p = 0.002$, partial $\eta^2 = 0.285$. The increase was 0.333, going from semi-transparent rendering with the mean accuracy of (4.733 ± 0.150) to opaque with accuracy ($5.107 \pm$

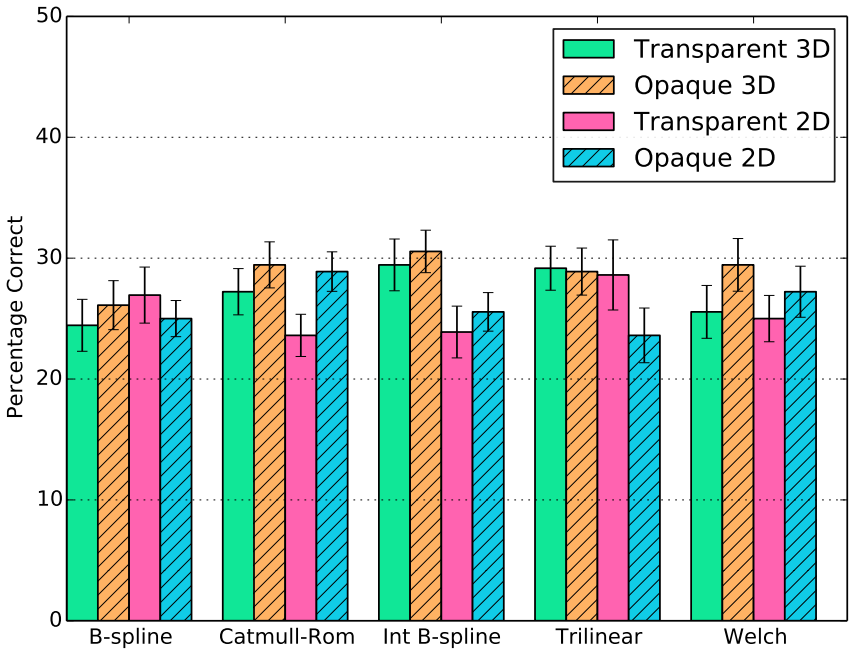


(a) Response accuracy

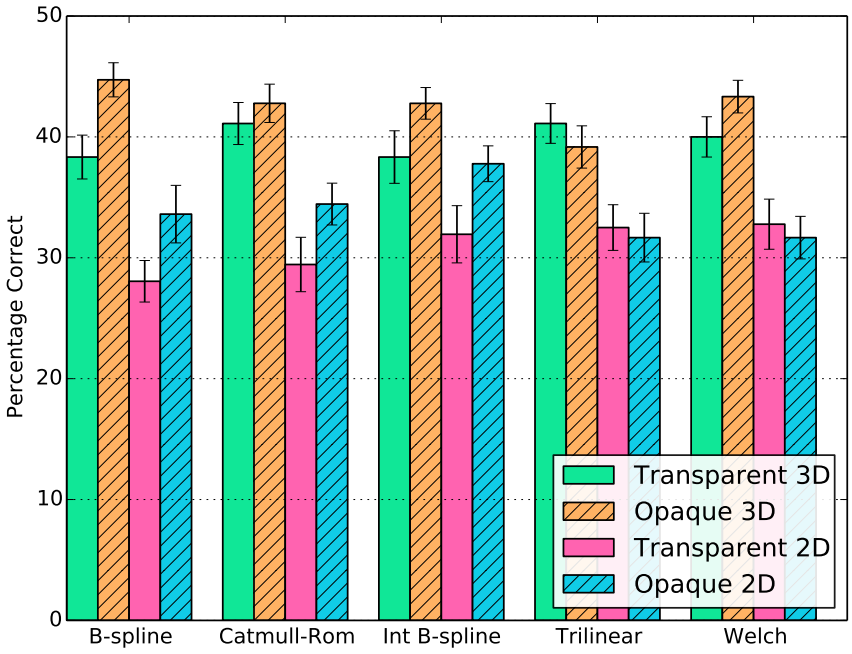


(b) Response time

Figure 6.5: Overall task performance with standard errors.



(a) Hits accuracy (true-positive)



(b) Correct rejections accuracy (true-negative)

Figure 6.6: Categorised accuracy rates with standard errors.

0.130).

A further significant effect was found when assessing the interactions between the opacity parameter and the reconstruction filter on the correct rejection rate. Statistical significance was found with $F(4, 116) = 2.985$, $p = 0.022$, partial $\eta^2 = 0.093$. Following this, simple main effects were run. Accuracy was significantly different between the semi-transparent rendering (4.600 ± 0.218) compared to the opaque rendering (5.367 ± 0.169) for the B-spline approximation filter, $F(1, 29) = 15.326$, $p = 0.01$, partial $\eta^2 = 0.346$.

The display had no significant effect on the number of Hits, however, it had a significant effect on the correct rejections with $F(1, 29) = 26.464$, $p < 0.001$, partial $\eta^2 = 0.477$ increasing from 2D (3.887 ± 0.131) to 3D (4.940 ± 0.131).

Response Time

Statistical significance was found only on the response times of the misses. In 3D, statistical significance was found with $F(3.126, 215.668) = 2.681$, $p = 0.046$, partial $\eta^2 = 0.037$. Post-hoc analysis revealed that there was no significant differences between individual pairs of reconstruction filters. In 2D, the reconstruction filter was also found to have a significant effect on misses with $F(2.599, 111.760) = 3.720$, $p = 0.018$, partial $\eta^2 = 0.103$. Post-hoc analysis revealed no significant pair-wise interactions.

6.3.3 Signal Detection Theory Analysis

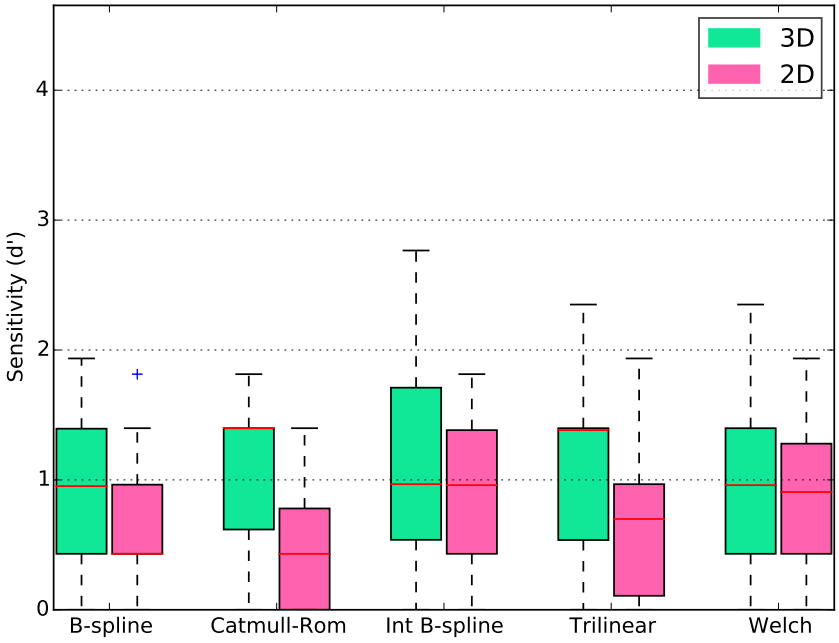
Sensitivity Measure

From the accuracy scores obtained, the sensitivity measure for each participant can be calculated. A high sensitivity means that participants had a good ability to detect the path, whereas a low sensitivity value means participants had a poor ability. The d' measure used to calculate the sensitivity, is defined as (Macmillan & Douglas Creelman 2004):

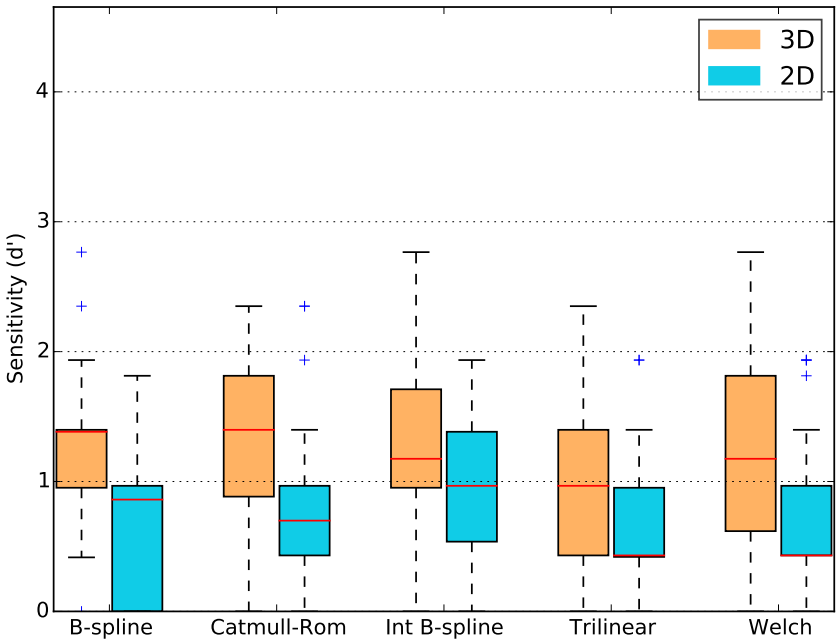
$$d' = z(H) - z(F) \tag{6.3.4}$$

where H is the conditional probability of a hit and F is the conditional probability of a false-alarm. $z(H)$ and $z(F)$ convert the hit and false-alarm probability rates to z-scores that are standard deviation units and have a range of $[0, 4.653]$.

Figure 6.7a and Figure 6.7b show the boxplots for d' for semi-transparent and opaque renderings respectively.



(a) Semi-transparent transfer function



(b) Opaque transfer function

Figure 6.7: Sensitivity measure for both opacity parameters tested.

In 3D, the opacity parameter had a significant effect on sensitivity with the results being $F(1, 29) = 4.650$, $p = 0.039$, partial $\eta^2 = 0.138$. Further significance was found when analysing interactions between the independent variables. The results of the ANOVA were $F(4, 116) = 2.487$, $p = 0.047$, partial $\eta^2 = 0.079$ and following this simple main effects were run. Significant results were found for the B-spline filter between semi-transparent rendering (0.879 ± 0.113) compared to the opaque rendering (1.252 ± 0.111) with $F(1, 29) = 9.034$, $p = 0.005$, partial $\eta^2 = 0.238$. Significance was also found for the Welch windowed sinc filter between semi-transparent rendering (0.935 ± 0.121) and opaque (1.272 ± 0.137) with $F(1, 29) = 5.046$, $p = 0.032$, partial $\eta^2 = 0.148$.

The display had a significant effect on sensitivity with $F(1, 29) = 42.216$, $p < 0.001$, partial $\eta^2 = 0.593$ increasing from 2D (0.731 ± 0.026) to 3D (1.150 ± 0.059).

Bias Response Measure

There was an overall preference towards a participant deciding that there is no connection between two nodes with the mean of No answers being 59.862%. The response bias can be measured using the Criterion Location measurement, defined in Macmillan & Douglas Creelman (2004) as:

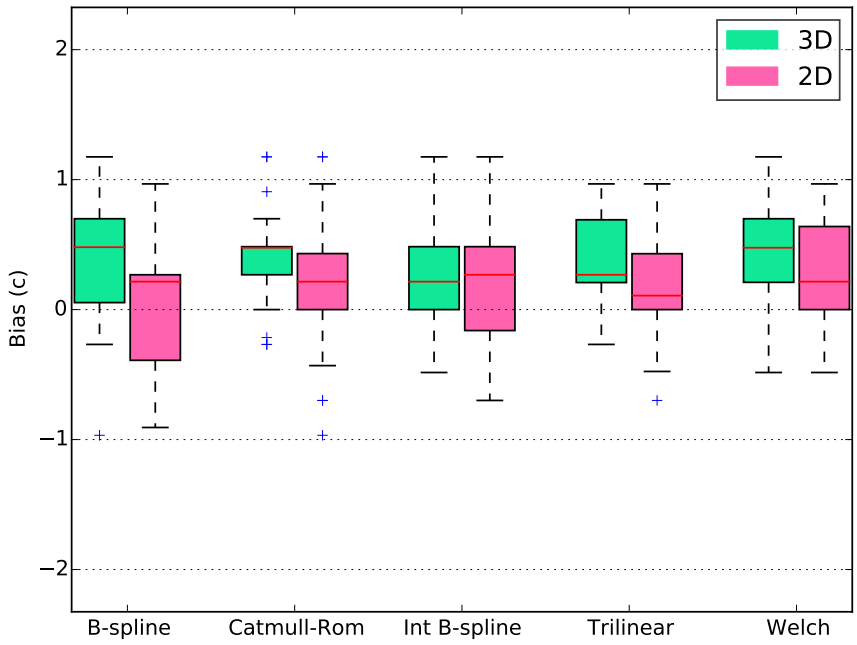
$$c = -\frac{1}{2}[z(H) + z(F)] \tag{6.3.5}$$

where H, F and z are as Equation 6.3.4 and the range of c is ± 2.325. A value of 0 is the neutral point and reflects an ideal participant who prefers neither response. A negative value reflects a *liberal* participant who is more likely to say yes. A positive value reflects a *conservative* participant who is more likely to err towards saying no. Figure 6.8a and Figure 6.8b show the boxplots for c for semi-transparent and opaque renderings respectively.

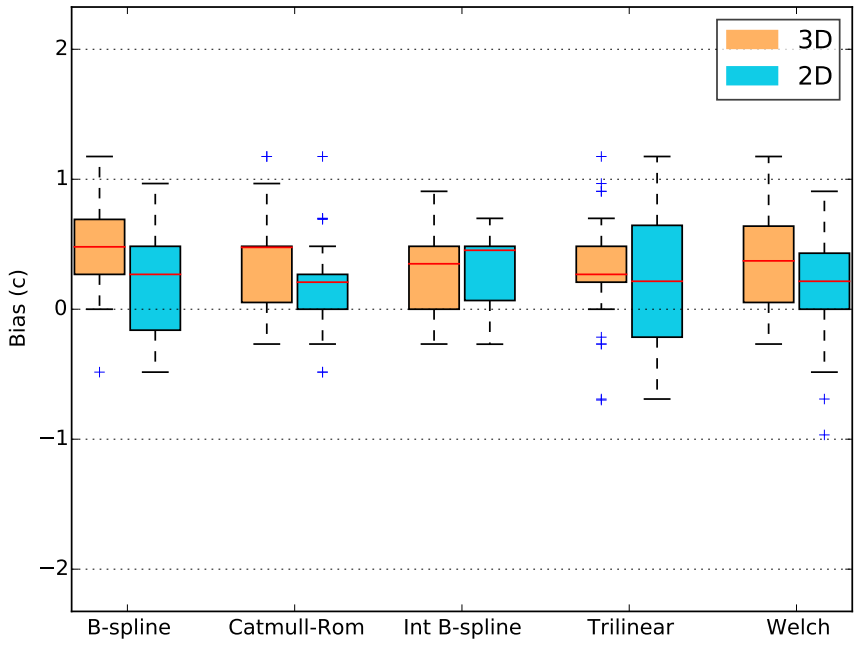
Only the display had a significant effect on bias, with the bias response measure increasing from 2D (0.191 ± 0.052) to 3D (0.371 ± 0.042) with $F(1, 29) = 7.670$, $p = 0.010$, partial $\eta^2 = 0.209$.

6.3.4 Correlation with Frequency Domain Metrics

A Pearson's product-moment correlation was run to assess the relationship between the post-aliasing properties of the reconstruction filters tested, the opacity and the reconstruction filter used to generate the images on the average accuracy of the conditions. Initial analyses showed the results to be linear with the variables normally distributed according



(a) Semi-transparent transfer function



(b) Opaque transfer function

Figure 6.8: Bias response measure for both opacity parameters used.

Table 6.3: Pearson Correlations for Accuracy

		Transparent 3D	Opaque 3D	Transparent 2D	Opaque 2D
Smoothing	R	-0.249	-0.744	0.099	-0.553
	Sig (p)	0.686	0.150	0.874	0.334
Post-aliasing	R	0.078	0.376	0.122	0.034
	Sig (p)	0.900	0.533	0.846	0.957

to the Shapiro-Wilk test ($p > 0.05$). No significantly statistical correlation was found between the variables and the post-aliasing property, with the full results being displayed in Table 6.3.

A further Pearson's product-moment correlation was run to assess the relationship between the smoothing property of each reconstruction filter, the opacity and the reconstruction filter used. Initial analyses showed the results were linear and the Shapiro-Wilk test determined the results were normally distributed ($p > 0.05$). There was no strong correlations found between the smoothing property and the variables, the full results are available in Table 6.3.

6.4 Discussion

6.4.1 Overall Accuracy and Response Time

Firstly, we discuss the overall results and compare to other studies. Due to other studies only reporting the overall accuracy and not hits or correct rejections, we are limited to just comparing findings on overall accuracy values and response times.

In 3D, the average accuracy rate of 69.195% is below that of similar, but not DVR, path-searching experiments with comparable complexities with approximately 85% being reported in Ware & Franck (1996), 90% in Ware & Mitchell (2005) and more recently 88.245% in Hassaine et al. (2010). The results are however within the range of depth discrimination tasks using stereoscopic DVR with angiography-like datasets. Cho et al. (2014) reports 67.59%, Ropinski et al. (2006) reports approximately 72% and Kersten-Oertel et al. (2014) reports 64.8%. The differences suggest that surface-base renderings, as used in the previous path-searching tasks, produce less ambiguity and therefore the use of low opacity must be carefully considered when using DVR. This result follows that of

Johnson et al. (2003) where stereoscopic displays combined with transparent surfaces led to ambiguous depth perception when viewing anatomical data. It also provides evidence for the results of Laha et al. (2014) where opaque iso-surface rendering was found to benefit from stereoscopic displays. However, it must be noted that surface rendering discards a significant amount of the volume data and precludes the ability to see through layers in the volume. Therefore a balance must be made between minimising ambiguity whilst still having the benefits of using DVR with semi-transparent rendering.

Contrary to the hypothesis, and previous research (Roberts et al. 2014, Mitchell & Netravali 1988), the choice of reconstruction filter was found to not have any impact on raw task accuracy for path-searching tasks. Instead, we find a more complex relationship between smoothing and post-aliasing on the one hand and participant responses on the other. We suggest that the differences between the results found in this experiment and the earlier experiment can be attributed to the introduction of noise and randomised line diameters used when generating the data sets.

Regarding the display, we found that there was a higher task accuracy when the stimulus was presented stereoscopically. That aligns well with results from previous studies (Kersten et al. 2006, Cho et al. 2014, Laha et al. 2012).

6.4.2 Hits and Correct Rejections

While the transfer function had a significant effect on correct rejections, increasing with opacity for both 2D and 3D, there was no significant effect on hits.

In 3D, an interaction effect was found between the level of opacity and the reconstruction filter on correct rejections for the B-spline and interpolating B-spline filters with low accuracy values found when combined with semi-transparent rendering. According to the Marschner and Lobb metrics, the filters have considerably different smoothing properties, yet their degrees of post-aliasing is comparable. It may then be suggested that low post-aliasing combined with semi-transparent rendering can lead to a decrease in the accuracy when assessing if two nodes are not connected.

Regarding the display, we found a significant effect on correct rejections but not on hits, meaning that the overall higher accuracy rates of the 3D display can be attributed to the higher number of correct rejections. In turn, the higher number of correct rejections in 3D can be partially attributed to the higher bias, that is, to the higher number of No answers given by the participants.

6.4.3 Signal Detection Theory

When analysing the sensitivity of participants to finding a path the opaque transfer function produced the highest sensitivity values. In 3D an interaction effect was also found between the opacity and the reconstruction filter with the B-spline and Welch windowed sinc filters producing the lowest sensitivity values. From the frequency domain metrics, the B-spline has the highest smoothing and Welch windowed sinc has the highest post-aliasing property. From this we can suggest that the sensitivity of participants to being able to detect the presence of a path is lowest when there is either a high amount of smoothing or post-aliasing.

From the bias response measure it was found that in 3D the reconstruction filter had a weak effect with the interpolating B-spline producing the least conservative participants and the B-spline producing the most conservative responses. From Figure 6.2 we can see that the B-spline filter has a considerable amount of smoothing but similar post-aliasing to the interpolating B-spline filter. We can therefore suggest that a filter that exhibits a large amount of smoothing with minimal post-aliasing may make participants more likely to be conservative for this type of task.

Regarding the display, we found that it had a significant effect on both sensitivity and bias. Notice that the simultaneous increase of overall accuracy, sensitivity and bias is a well understood phenomenon related to the shape of the ROC curves the participants operate on, see Macmillan & Douglas Creelman (2004).

6.5 Conclusions

This chapter has evaluated the effect of five reconstruction filters and two opacity levels on a path-searching task with stereoscopic and monoscopic DVR. We found that there was a considerable amount of conservative bias in the responses of the participants, which then was taken into account in the further analysis of the results.

In the future, a meta-analysis of previously published papers on the effect of stereoscopic visualisation on path-searching tasks should be performed. Even though the experimental results of such path-searching tasks are customarily reported in a way that makes bias computations impossible, nevertheless, approximations of the most common sensitivity measure d' are possible (Macmillan & Douglas Creelman 2004) and could yield new interesting insights to their results.

Chapter 7

Quality Measures of Reconstruction Filters for Stereoscopic Volume Rendering

This chapter is based on work published in 'Quality Measures of Reconstruction Filters for Stereoscopic Volume Rendering' (Roberts & Ivrissimtzis 2016*a*). In it, we move from objective measured task performance to eliciting subjective responses from users. The chapter evaluates four quality measures on different stereoscopic DVR scenes and analyses the scores in terms of frequency domain properties and shape compactness.

7.1 Introduction

The previous chapters have found that the choice of reconstruction filter can affect the results of path-searching tasks when used with stereoscopic DVR. However, it is not clear how observers perceive the different artefacts produced by the filters, or how they may affect subjective measures of the quality of a stereoscopically viewed DVR rendered scene.

This chapter presents an experimental study of four quality measures related to stereoscopic volume rendering: depth quality, depth layout, minimal jaggyness and sharpness. We tested five reconstruction filters on four different data sets, three of which are natural data sets and one is computer generated. We then analysed the results looking for relationships between the quality measure scores gathered from the participants on the one hand and the smoothing and post-aliasing of the filters and the shape compactness of the images on the other.

The analysis of the results of the experiment suggests that while the perception of the quality of the scene layout is not affected by the choice of reconstruction filter, depth quality is. That is, whilst users are able to understand the global layout of objects in the scene regardless of the choice of reconstruction filter, the perception of depth quality depends on that choice. However, when we tried to translate the statistically significant relationships between quality measures filters into correlations between quality measures and properties of the filters, here smoothing and post-aliasing, we found such correlations to be moderate at best and statistically insignificant. On the other hand, we found that the image specific, rather than filter specific, measure of shape compactness is highly correlated with each of the quality measures.

The implications of the results is firstly that the often quoted smoothing and post-aliasing measures of reconstruction filters for DVR do not necessarily relate to how well observers will perceive the images. Therefore preferences towards reconstruction filters that exhibit high or low post-aliasing or smoothing do not guarantee any particular level of perceived quality. Measures that take into account the rendered images, such as the shape compactness used in this chapter, should be preferred instead. Secondly, the shape compactness results suggest that the B-spline approximation filter should be preferred if the smoothing of the volumetric data caused by the use of an approximating filter can be tolerated.

7.1.1 Quality of Stereoscopic DVR

The quality of stereoscopically viewed DVR has been primarily assessed using tasks that require a participant to perform a task related to depth perception. These include the depth ordering evaluations of Boucheny et al. (2009), Cho et al. (2014) where volume rendered cylinders are displayed and participants must find their ordering. As well as the more common relative depth tasks requiring a participant to find which highlighted feature is closest to themselves (Ropinski et al. 2006, Kersten-Oertel et al. 2014) and path-searching tasks (Agus et al. 2009, Roberts et al. 2014). These studies use the task's accuracy rates to compare the quality of one rendering method or display system over another but do not give any information as to how participants perceive any visual differences.

The literature on the assessment of the perceived quality of stereoscopic DVR is more sparse. In Prabhat et al. (2008) an open comment system was used for participants to remark upon the quality of different display systems used for DVR. Participants positively

commented on the quality of a Fishtank display versus a standard 2D display and gave positive comments towards the high levels of immersion in the Cave environment. More formalised quality measures were used in Laha et al. (2012) with a Likert scale being used to determine the comfort level, ease of viewing and ease of understanding features in the data sets. The results found that the ease of viewing increased for stereoscopic viewing. The qualitative measures used in these studies are however secondary to the primary task based evaluations performed.

7.1.2 Shape Compactness

Although there are several methods for defining compactness for planar shapes, the measure used in Li et al. (2013) is the most widely accepted:

$$C = \frac{4\pi A}{P^2} \tag{7.1.1}$$

where A is the area of the shape and P its perimeter. Because of the isoperimetric inequality (Osserman 1978), the compactness value of a shape is always in the range $(0, 1]$. The highest value of compactness is equal to 1 and attained by the disc, while shapes with multiple branching extremities have low values of compactness that approach 0. It is scale invariant and moreover is relatively stable under the almost rigid shape deformations that are common in physical simulations.

Implementation Details

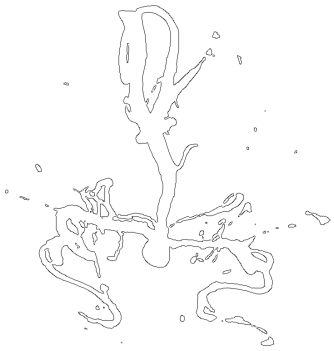
A simple method for calculating a compactness value for a DVR image is by computing the area and perimeter of the features of the image. By applying a binary thresholding to the DVR image a silhouette of the features is produced. The area of the features is then calculated by counting the number of non-zero pixels while the perimeter is found by applying the Canny edge detection algorithm to the silhouette and counting the number of edge pixels.

We considered two methods of binary thresholding: manual fine tuning for each data set and Otsu's method (Otsu 1979). Figure 7.1 displays the results of the two thresholding methods and the corresponding edge sets returned by the Canny algorithm. Otsu's methods can be seen to be more conservative than manual thresholding. In our analysis we always used Otsu's method since it is fully automatic and thus it was deemed more objective.

(a) Manual thresholding

(b) Canny on manual thresholding

(c) Otsu Thresholding



(d) Canny on Otsu thresholding

Figure 7.1: Thresholding of an angiography data set using the two methods considered.

Obviously, the compactness values of the DVR images depend heavily on the data set. For example, the engine data set would most likely produce more compact images than the angiography data set, under any reasonable choices of parameter values. On the other hand, and critically for our purposes, the choice of reconstruction filter can also significantly alter the compactness of the rendered images. For example, a filter with a significant smoothing effect can blur and eventually remove from the thresholded image the thinner branches, increasing the overall compactness. It should be noted that this implementation does not take into account partial volume effects where a sample point may contain values from different materials.

7.2 Experiment Methodology

7.2.1 Quality Measures

In the assessment of the quality of the stereoscopic volume rendered images and the properties of each reconstruction filter, four different quality measures were used. Table 7.1 shows the descriptions of the quality measures as given to the participants.

The first, depth quality, is a critical aspect of stereoscopic viewing and is dependant upon two factors: the stereoscopic image being presented as well as the hardware and viewing settings used to display it. Artefacts such as aliasing and blurring in the image can cause errors between perspective and binocular depth cues (Pfautz 2000) and degradation of fine depth perception (Costa et al. 2010) that may lead to poor depth quality. Low display resolutions and incorrect viewing distances are typical hardware and viewing setting factors that can lead to a poor quality of depth perception.

To determine how reconstruction filters may affect the ability to understand the organisation of the features of the scene, the second measure that was assessed is the depth layout (Lebreton et al. 2012). It characterises how positions and spatial features in a scene are understood. While the depth effect and perspective cues are considered more helpful than stereopsis when determining object positions and how they relate to each other (Ware & Franck 1996), nevertheless, when sampling artefacts exist, stereopsis combined with linear perspective can provide conflicting depth information to the viewer (Pfautz 2000).

The third quality measure was sharpness, which concerns the crispness of details in the stereoscopic images presented. Previous subjective evaluation of stereoscopic images

Table 7.1: Explanations of quality measures given to participants

Quality Measure	Explanation
Depth Quality	Is the depth presented in the scene realistic
Depth Layout	How easy is it to understand the organisation of the elements composing the scene
Sharpness	Clarity of the details in the image
Minimal Jagginess	Excellent if there are no jaggies or staircase artefacts, bad if there are a high number

have shown no strong correlations between sharpness and depth quality, necessitating the need for the sharpness and an independent measure (Tam 1998, Seuntiens et al. 2006).

The fourth assessed quality measure assessed was minimal jaggyness. In this case low values of that measure would reflect no jaggy or aliasing artefacts. Increased levels of aliasing in stereoscopic images have been suggested to cause errors in disparity and conflicting depth information (Pfautz 2000). Further, it has been demonstrated that anti-aliasing techniques can increase the ability to perceive depth (Jää-Aro & Kjelldahl 1997). One would reasonably expect that reconstruction filters with higher levels of aliasing would lead to higher jagginess scores, however that had to be tested since it has been suggested that a form of binocular averaging occurs between the two eyes that decreases the overall perceived aliasing artefacts (Hancock 2001, Reinhart 1992).

7.2.2 Reconstruction Filters

Five reconstruction filters have been evaluated: B-spline, trilinear, Catmull-Rom, interpolating B-spline and the Welch windowed sinc. The B-spline filter is an approximation filter and technically not a reconstruction filter, but it is included due to its common usage for noisy data sets. The trilinear filter, being built into GPUs, is the most commonly used filter in DVR. The Catmull-Rom is an interpolating spline with a balance between smoothing and post-aliasing artifacts. The interpolating B-spline uses a prefilter to force the B-spline to interpolate the values at the grid nodes, at the cost of increased post-aliasing artefacts. Finally, the Welch windowed sinc is a windowed version of the sinc filter that has minimal smoothing but high post-aliasing.

The frequency domain smoothing and post-aliasing metrics of (Marschner & Lobb 1994) measure the pass-band and stop-band behaviour, respectively. Figure 6.2 in Chapter 6 shows the values of the smoothing and post-aliasing metrics for the five reconstruction filters used in the experiment.

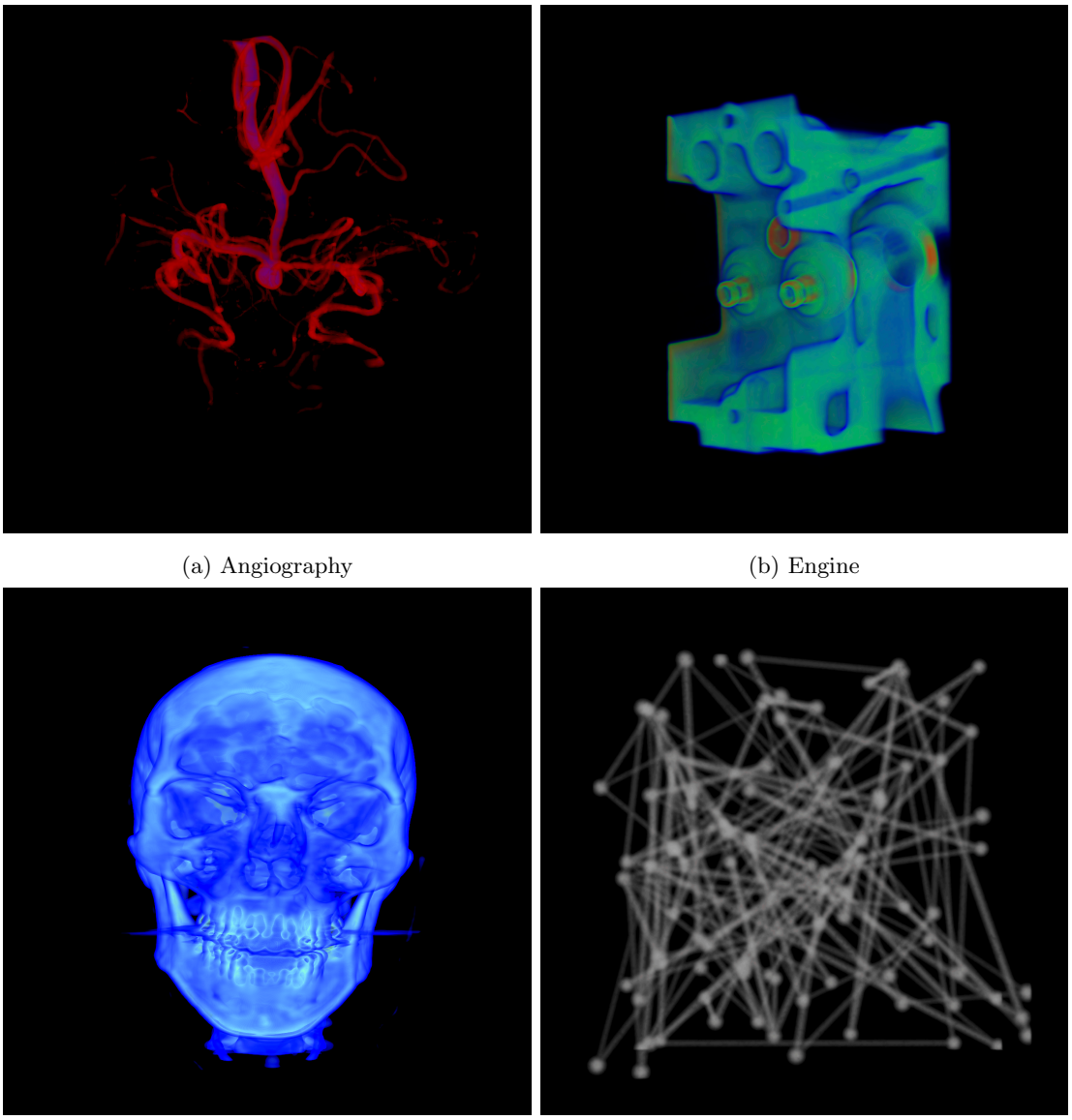
7.2.3 Stimuli

Four data sets were used in the experiment. Three of them, the angiography, the engine, and the Visible Human are real and have been created via conventional scanning processes. The volume resolutions of these data sets are $(512 \times 512 \times 512)$, $(256 \times 256 \times 256)$ and $(128 \times 256 \times 256)$ respectively. For each data set the transfer function that assigns colours and opacities has been chosen arbitrarily. Figures 7.2a, 7.2b and 7.2c show renderings of each of these data sets using the B-spline reconstruction filter.

The fourth data set, referred to as the graph, is a computer generated data set containing a set of spherical nodes connected to other nodes via cylinders. It has been designed to encapsulate characteristics of natural data sets such as noise, branching and complex spatial layout. The noise has been introduced by modulating the volume data with a 3D Perlin noise function and the spatial structure of the graph is determined by a graph layout algorithm detailed in Ware & Franck (1996). The use of noise and graph layout algorithms produces a volume approximating the irregularities of the natural data sets whilst allowing the overall complexity to be easily controlled by increasing or decreasing the number of nodes. The data set was produced by the publicly available software *vxtRL* (Novotny et al. 2010) at a resolution of $(256 \times 256 \times 256)$. The transfer function was designed to have a degree of semi-transparency whilst still displaying clear structures. Figure 7.2d shows a rendering of the graph data set with the B-spline filter.

7.2.4 Equipment and Viewing Conditions

To display the 3D images a stereoscopic True3DI 24" HD-SDI display was used with a resolution of 1920×1200 for each eye and refresh rate of $60Hz$. A 17 inch IBM LCD display, used to display the grading scales for the measurement scores, was positioned to the left of the stereoscopic display. Participants wore linear polarized glasses at all times during the experiments and were seated at a distance of $60cm$ from the display. All light levels were kept constant throughout the experiments.



(a) Angiography (b) Engine

(c) Visible Human (d) Graph

Figure 7.2: Example renderings of the four data sets used in the experiment.

7.2.5 Participants

Overall, 20 participants took part in the experiment; 14 women and 6 men with ages ranging from 18 to 29. All participants were paid £5 for taking part in the experiment. The Bailey and Snellen chart were used to screen participants for vision and the Titmus fly stereo test was used to screen for stereo-vision.

The participants used for the experiment did not have experience with analysing volumetric data sets, however some were familiar with stereoscopic displays. Although domain experts can provide highly specific results, novice participants are free of preconceived biases towards particular display methods or rendering styles.

7.2.6 Procedure

Before the main experiment, each participant took a pretrial test where five images were displayed and rated for each of the four measurements. For the pretrial test no scores were recorded. In the main experiment, for each participant, the images of each data set were shown as a block sequence, i.e. there was no interleaving between data sets. The ordering of the four data sets for each participant was determined via a Latin Square design, while the ordering of the reconstruction filters for each data set was random.

For each image in the experiment four grading scales were used, one for each of the quality measurements assessed. The sliding scales were divided into five regions labelled Excellent, Good, Fair, Poor and Bad according to the ITU-R BT.500 recommendations (ITU-R 2002). The scale was continuous with values ranging from 0-100, which were not presented to the participant. After submitting the scores for an image, a blank screen was displayed for 3 seconds. Figure 7.3 shows the grading scales used for the experiment.

7.3 Results

Section 7.3.1 presents the statistical analysis of the results of the experiment for the Angiography, Engine, Visible Human and Graph data sets, respectively. In Section 7.3.2, we analyse the relationship between the smoothing and post-aliasing of the reconstruction filters and the quality scores. Finally, in Section 7.3.3 we analyse the relationship between the compactness of the test images and the quality scores.

Figure 7.3: Grading scales used for the subjective experiment

7.3.1 Results of the Experiment

Prior to performing statistical analysis of the results, the subjective scores from each participant for each measure were transformed to z-scores. This is required because not all participants use the entire range of the scales presented when rating the images, therefore their scores are not directly comparable. To normalize the raw scores into z-scores the values are converted into standard deviation units that represent deviations from the mean (van Dijk et al. 1995). Figure 7.4 shows the average z-score values for each data set, quality measure and reconstruction filter.

Following the ITU-R BT.500 recommendations, an analysis of the results per individual participant was performed prior to the main statistical analysis (ITU-R 2002). This was done to potentially remove participants who may have misunderstood the individual quality measures used in the experiment. The β^2 normality test was used for each participant and quality measure separately. The test determined that no participants should be removed prior to full analysis.

For the main analysis, the z-scores for each measure per data set were subjected to a one-way repeated measures of analysis of variance (ANOVA) with the reconstruction filter as the independent variable and the score as the dependent variable. In the cases when the assumption of sphericity failed after applying Mauchly's Test of Sphericity, the Greenhouse-Geisser correction was applied. When applicable, post-hoc analysis was

performed using the pairwise Bonferroni test.

Angiography

The choice of reconstruction filter was found to have a significant effect on depth quality when viewing the angiography data set. The results of the ANOVA were $F(4, 76) = 3.152$, $p = 0.019$, partial $\eta^2 = 0.142$. Despite the interpolating B-spline having the lowest depth quality with a z-score of -0.253 and the Welch windowed sinc having the highest with 0.473 a post-hoc pairwise Bonferroni test revealed no significant interactions between individual pairs of reconstruction filters.

Regarding the sharpness measure, the choice of reconstruction filter was found to have a significant effect with the results of the ANOVA being $F(4, 76) = 2.991$, $p = 0.024$, partial $\eta^2 = 0.071$. Following a post-hoc Bonferroni test, significant differences were found between the trilinear (0.362) and interpolating B-spline (-0.433), $p = 0.044$.

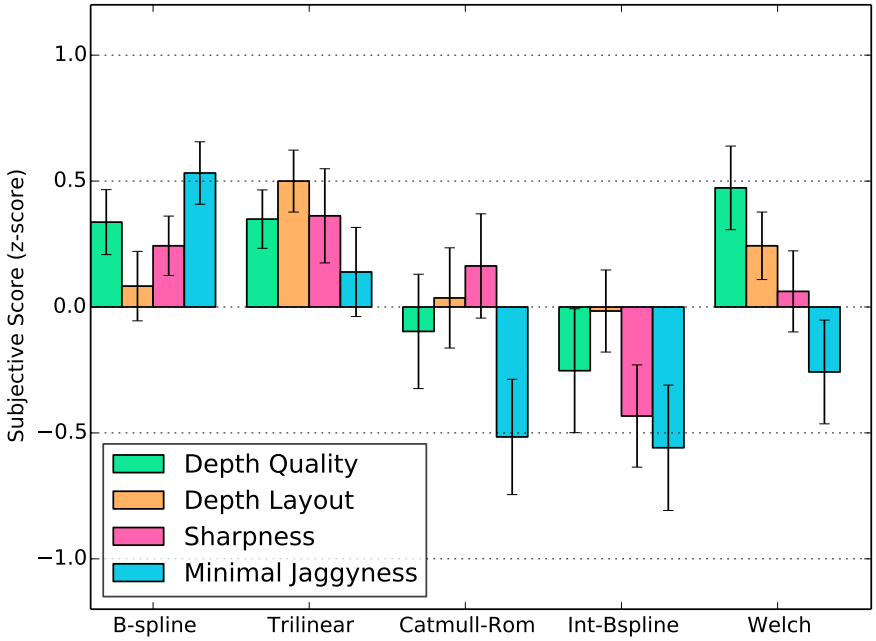
Significant differences were found when analysing the scores for the minimal jaggyness measure with $F(4, 76) = 5.739$, $p < 0.001$ and partial $\eta^2 = 0.232$. According to a post-hoc Bonferroni pairwise test, the B-spline (0.532) was significantly different from the Catmull-Rom (-0.516), $p = 0.008$ and the interpolating B-spline, (-0.559), $p = 0.003$.

There was no statistical significance of reconstruction filters on the depth layout measure for the angiography data set.

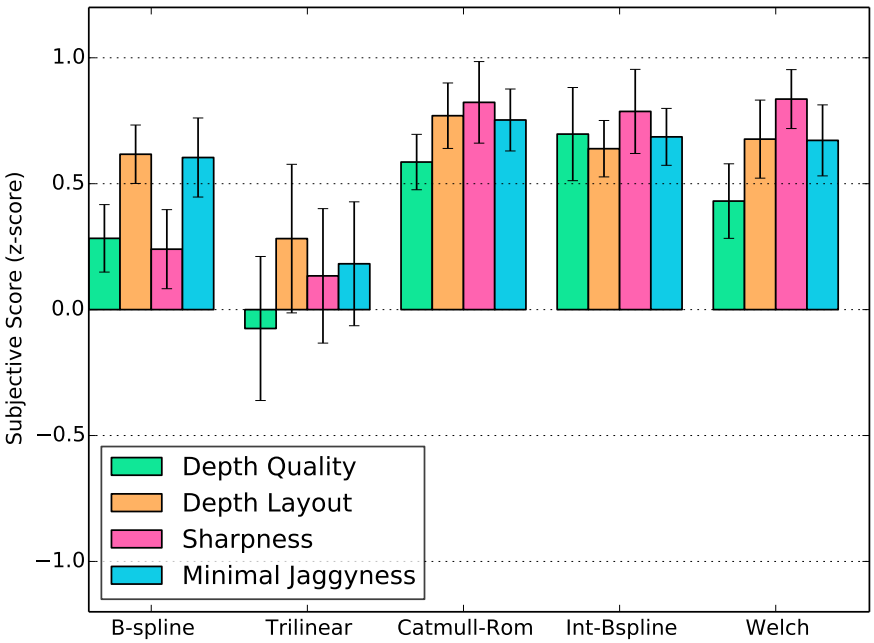
Engine

For the engine data set, the choice of reconstruction filter was found to have a significant effect on depth quality. Mauchly's test of Sphericity was violated with $p = 0.005$ and therefore the Greenhouse-Geisser correction was used. The results of the ANOVA were $F(2.492, 47.537) = 3.284$, $p = 0.036$, partial $\eta^2 = 0.147$. No significant interactions between pairs of reconstruction filters were found according to a post-hoc pairwise Bonferroni test. However, the trilinear filter had the lowest depth quality with an average of -0.075 and the interpolating B-spline had the highest with 0.697.

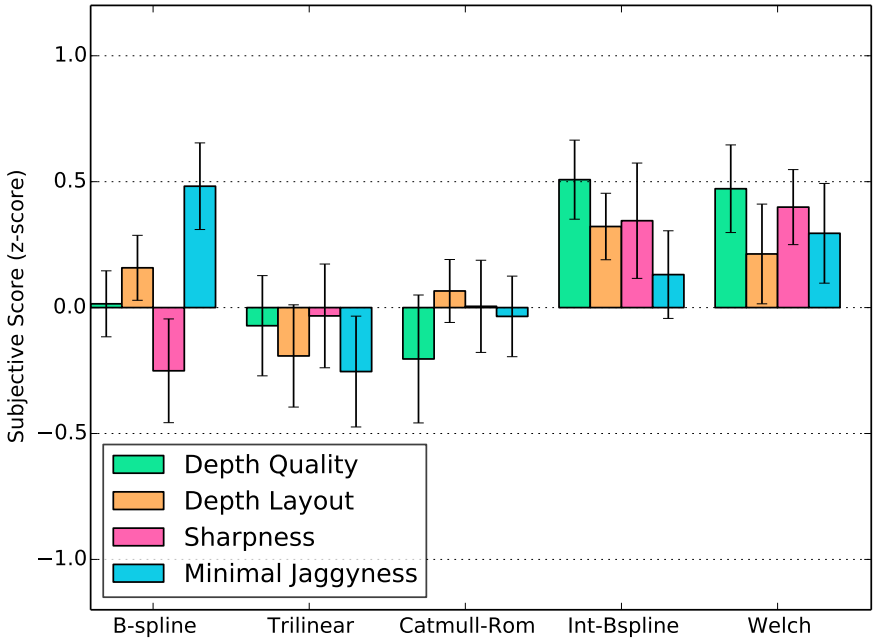
The choice of reconstruction filter was found to have a significant effect on sharpness, with the results of the ANOVA being $F(4, 76) = 4.582$, $p = 0.002$, partial $\eta^2 = 0.194$. A follow-up Bonferroni test reported that the sharpness score of the B-spline filter (0.240) was significantly lower than the interpolating B-spline (0.787), $p = 0.025$ and the Welch windowed sinc (0.836), $p = 0.016$.



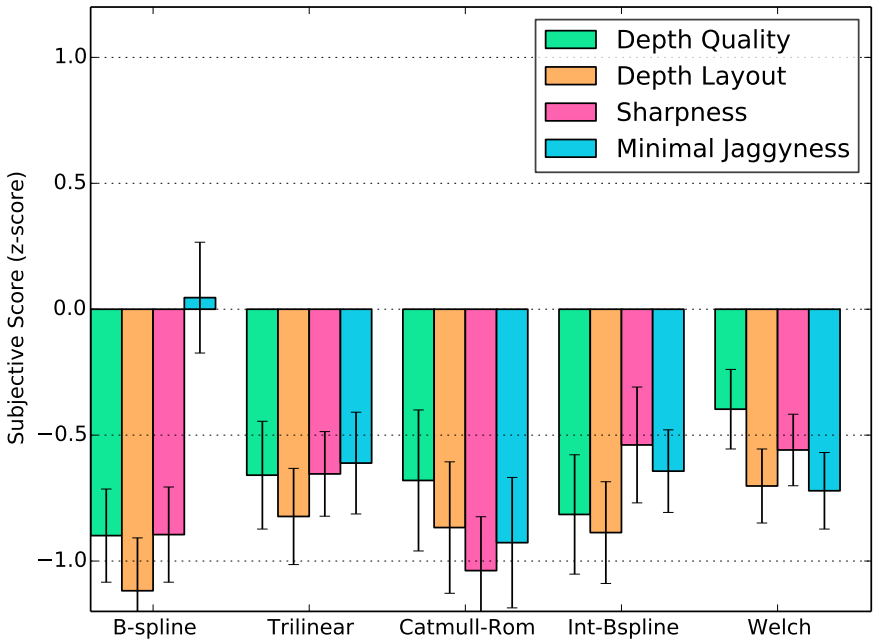
(a) Angiography



(b) Engine



(c) Visible Human



(d) Graph

Figure 7.4: Plot of average z-score results with standard errors.

The results of ANOVA showed that there was no significant effect of the choice of reconstruction filter on depth layout or minimal jaggyness for the engine data set.

Visible Human Data Set

A significant effect was found on depth quality with the ANOVA reporting $F(4, 76) = 4.612$, $p = 0.002$, partial $\eta^2 = 0.195$. Follow-up post-hoc tests revealed no significant interactions between pairs of reconstruction filters, however the Catmull-Rom filter had the lowest average with -0.204 and the interpolating B-spline had the highest with 0.508.

The ANOVA for minimal jaggyness also reported statistical significance. Sphericity had been violated with $p = 0.005$ and so the Greenhouse-Geisser correction was applied. The ANOVA results were $F(2.678, 50.883) = 3.039$, $p = 0.043$, partial $\eta^2 = 0.138$. No interactions were found between individual reconstruction filters, however the trilinear filter had the lowest average with -0.254 and the B-spline had the highest with 0.482.

The choice of reconstruction filter had no statistically significant effect on the depth layout and the sharpness scores.

Graph

The choice of reconstruction filter was found to only have an effect on the minimal jaggyness measure with the ANOVA reporting $F(4, 76) = 4.308$, $p = 0.003$. A Bonferroni test reported that the B-spline filter (-0.046) was significantly different from the Welch windowed sinc filter (-0.721) with $p = 0.013$.

No significant results were reported for the depth quality, depth layout and the sharpness measures.

7.3.2 Correlations with Smoothing and Post-aliasing

The smoothing and post-aliasing introduced by the reconstruction filter correspond to the blurring and aliasing of the features of the rendered images. These artefacts can degrade depth perception when viewing stereoscopic scenes (Costa et al. 2010, Pfautz 2000). Here, we investigate whether these properties of the filters are related to how observers perceived the images by measuring the correlation between the smoothing and post-aliasing metrics on the filters proposed in Marschner & Lobb (1994), and the quality scores obtained from the experiment.

Table 7.2: Correlations between smoothing and post-aliasing and quality scores

		Depth Quality	Depth Layout	Minimal Jaggyness	Sharpness
Post-aliasing	r	0.210	0.099	-0.166	0.191
	p	0.374	0.677	0.483	0.419
Smoothing	r	-0.192	-0.096	0.249	-0.194
	p	0.418	0.686	0.289	0.412

The Pearson correlation coefficients were computed to determine if a linear relationship exists between the smoothing and post-aliasing properties and the quality scores obtained from the experiment. A summary of the results can be seen in Table 7.2. There were no strong correlations between either of the frequency domain metrics and the quality scores obtained.

7.3.3 Correlations with Compactness

The frequency domain metrics used to measure smoothing and post-aliasing are filter specific in that they are independent of the data set used or images produced. As an alternative we investigated whether the image specific measure of shape compactness was correlated with the quality scores obtained.

The compactness of each image used in the experiment was computed by the method described in Section 7.1.2. We found that the compactness analysis gave very similar results when it is applied either on the left image of the stereoscopic pair, the right image, or on the combined cyclopean image. For simplicity, here we only report the analysis for the left images. The compactness measures for each of the twenty images used in the experiment can be seen in Figure 7.5a and Figure 7.5b. From the results we can see that the Engine and Visible Human data sets have high compactness values where as the Angiography and Graph data sets have low values. In all cases, the use of the B-spline approximation filter gave the highest compactness.

After computing the compactness measures for the images Pearson correlation coefficients were computed to see if a relationship exists between image compactness and the quality scores. A summary of the results is shown in Table 7.3. There was a moderate positive correlation between the compactness measure and the quality of depth reported by the participants, $r(18) = 0.475$, $p < 0.05$ with compactness statistically explaining

Table 7.3: Relations between compactness and quality scores. * denotes significance at 0.05, ** denotes significance at 0.01.

		Depth Quality	Depth Layout	Minimal Jaggyness	Sharpness
Compactness	r	0.475*	0.557*	0.605**	0.510*
	p	0.034	0.011	0.005	0.022

22.563% of the variation in depth quality. A strong correlation was found for depth layout with $r(18) = 0.557$, $p < 0.05$ with compactness explaining 31.025% of the variation in depth layout. Compactness was also found to have a strong correlation with minimal jaggyness with $r(18) = 0.605$, $p < 0.01$ and compactness statistically explaining 36.603% of the variation in scores found. A further strong correlation was found for sharpness with $r(18) = 0.510$, $p < 0.05$ and compactness explaining 26.010% of the variance in sharpness scores.

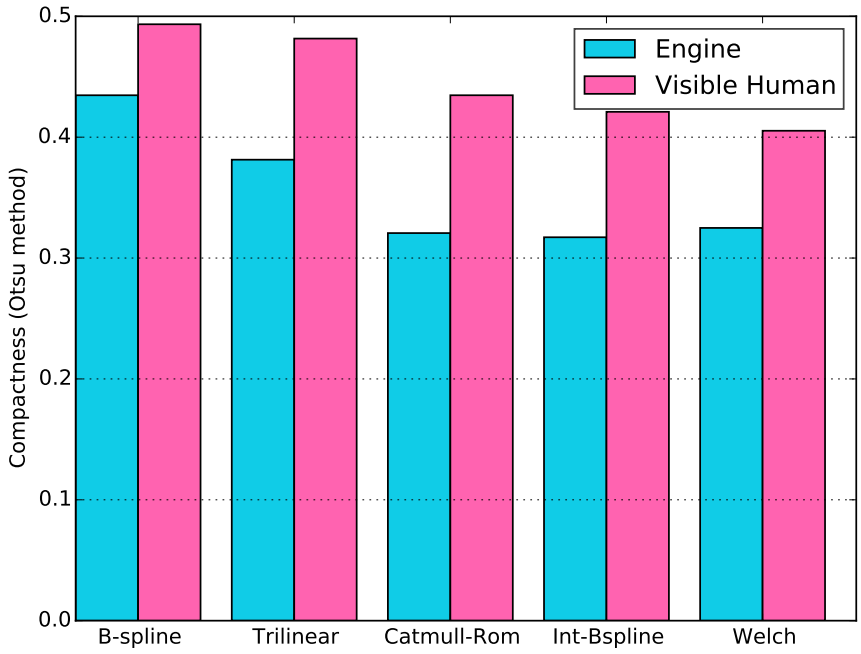
7.3.4 Correlations between Quality Measures

We calculated Pearson correlation coefficients in order to examine the relationships between depth quality, depth layout, minimal jaggyness and sharpness. A summary of the results can be seen in Table 7.4. The results found showed significant positive correlations between each pair of quality measures used in this experiment.

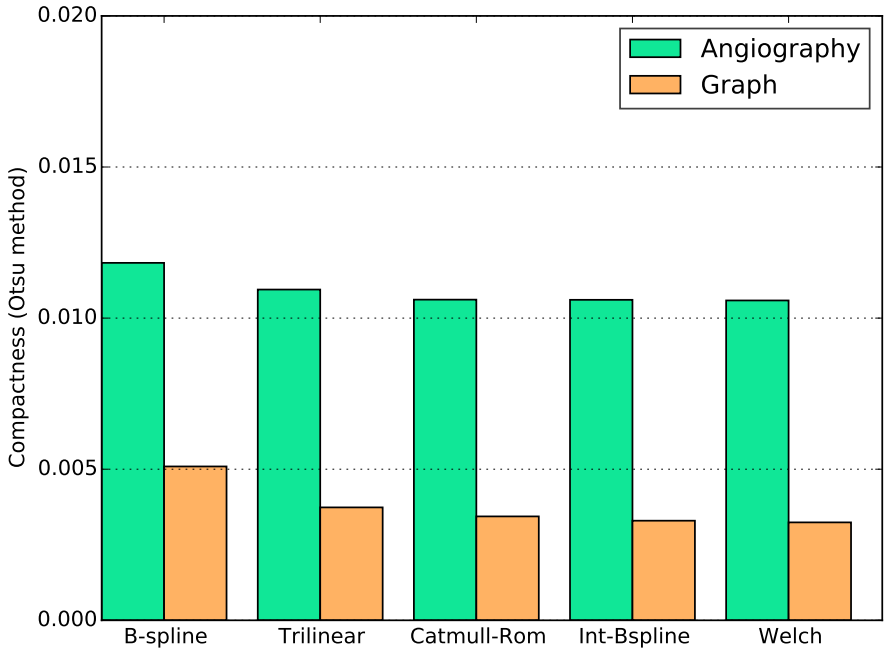
7.4 Discussion

From the overall analysis of the results, the largest factor in the differences in quality scores is the data set itself. The natural data sets reported higher scores for the quality measures than for the computer generated data set, implying a preference towards non-computer generated data sets.

However, in many practical situations users do not have control over the data sets they view and it is therefore a significant finding that the choice of reconstruction filter had an effect on the quality scores. Indeed, in the three natural data sets, the choice of reconstruction filter was found to have a significant effect on the depth quality. On the other hand, exactly how the choice of reconstruction filter affects depth quality is still not clear. The interpolating B-spline had the highest depth quality measure for the Engine



(a) Engine and Visible Human data sets. Range 0-0.5



(b) Angiography and Graph data sets. Range 0-0.02

Figure 7.5: Plots of compactness measures split into two graphs with different scalings of the y-axis due to the polarization of the results.

Table 7.4: Correlations between quality measures. * denotes significance at 0.05, ** denotes significance at 0.01.

		Depth Quality	Depth Layout	Minimal Jaggyness	Sharpness
Depth Quality	r	-	0.924**	0.736**	0.915**
	p	-	<0.001	<0.001	<0.001
Depth Layout	r	-	-	0.764**	0.922**
	p	-	-	<0.001	<0.001
Minimal Jaggyness	r	-	-	-	0.779**
	p	-	-	-	<0.001
Sharpness	r	-	-	-	-
	p	-	-	-	-

and the Visible Human data sets, while for the Angiography data set the interpolating B-spline had the lowest score. This suggests that some processing and analysis of the volumetric data is required before choosing the optimal reconstruction filter for viewing them.

Reconstruction filters were found to have no significant effect on the depth layout scores for any of the data sets used in the experiment. This implies that the differences in the images do not affect the general ability to understand global features in the scene. It also suggests that the perceived relationships between object positions are unaffected by the choice of reconstruction filter.

As expected from an approximating filter, the B-spline had significantly higher minimal jaggyness scores in most cases. This implies that it produces images with the fewest number of perceived aliasing artefacts, meaning that its low Marschner and Lobb post-aliasing value has visual significance.

For the Angiography and Engine data sets the sharpness quality measure was affected by the reconstruction filter, however the results vary considerably between the data sets. This suggests that the data set is the primary factor affecting the values of this measure.

7.4.1 Smoothing and Post-aliasing

The Pearson correlation coefficients indicate that the smoothing and post-aliasing properties of the reconstruction filters have only small correlations with the quality measures, i.e., r coefficients in or very near the region of $[0.1, 0.3]$. Moreover, these measured correlations are not statistically significant. This suggests that a choice of reconstruction filter based on its smoothing and post-aliasing properties alone does not guarantee any level of quality and that other data set specific factors must be considered.

7.4.2 Compactness

Unlike smoothing and post-aliasing, the shape compactness measure was found to have a strong correlation to each of the quality measures assessed in the experiment, i.e., r coefficients in or very near the region of $[0.5, 1.0]$. In each case, a high compactness measure implies a high score for depth quality, depth layout, minimal jaggyness and sharpness. If a choice of data set is possible, then data sets with high compactness, such as the Engine and the Visible Human, should be preferred. Of course, in many applications we do not choose the data set. In such cases, the compactness values in Figure 7.5 show that the B-spline filter produces the highest compactness across all data sets and so should be preferred. If an approximation filter is not suitable, then according to the results the trilinear filter should be the second preference

7.4.3 Quality Measures

Each quality measure used was found to have a strong correlation with each of the other quality measures with coefficients in the regions of $[0.73, 1.0]$ with each of the correlations being strongly statistically significant. The strong correlation between depth quality and sharpness is particularly interesting. Prior studies have found either only a low correlation or no correlation between these two measures for stereoscopic video sequences and compressed stereoscopic images respectively (Tam 1998, Seuntiens et al. 2006). Perhaps unsurprisingly depth quality and depth layout were found to have the strongest correlation. This is despite the choice of reconstruction filter only affecting depth quality showing that both measures are required.

7.5 Conclusion

In this chapter we have evaluated several quality measures on stereoscopic DVR images produced by various reconstruction filters. The results suggest that the choice of reconstruction filter affects the perceived quality of depth in stereoscopic DVR images but not the understanding of the layout of objects in the scene. We have found that the frequency domain measures of smoothing and post-aliasing of individual filters do not correlate with any of the quality measures. Instead, a positive correlation was found between the compactness of the images and each of the quality measures, with the B-spline approximation filter producing the most compact images.

From a practical point of view, the results suggest that we cannot choose an optimal reconstruction filter using the filter specific measures of smoothness and post-aliasing. However, based on the strong correlation between image compactness and depth perception, a possible strategy for choosing reconstruction filter would be to render the dataset with several filters, compute the compactness measure of each rendered image and, finally, present to the viewer the image with the highest compactness.

Chapter 8

Conclusions and Future Work

8.1 Introduction

This thesis has investigated stereoscopic direct volume rendering by evaluating the performance of tasks that require accurate depth perception and an understanding of how objects in the scene are organised. Further, the thesis has evaluated the perceived quality of stereoscopic DVR on four different factors and metrics that can be used to measure these. A particular focus has been on the choice of the parameters in the direct volume rendering pipeline including: the resolution of the volumetric data, the reconstruction filter used for generating sample points and the transfer function.

In this chapter the overall results and contributions of the thesis are summarised, directions for future work leading from the research are described and then a final conclusion is presented.

8.2 Research Contributions

The contributions of the research on stereoscopic volume rendering given in this thesis are summarised below.

1. **An evaluation of the relationship between volume resolution and fine depth perception on stereoscopic volume rendering.**
   In Chapter 4 an initial investigation into how parameters in the DVR pipeline can affect depth perception is detailed. Three volume resolutions and one reference rendering are evaluated with a stereo-acuity test. The results concluded that the volume resolution had no statistical significance on fine depth perception for the

resolutions tested. Statistically significant results were found for response latency with the results implying that higher volume resolutions can increase the time it takes to make a decision. We hypothesise that the smooth edges of the data sets disrupt the HVS from matching stereo images and therefore cause an increase in response latency.

2. **Investigation of the influence of choice of reconstruction filter on path-searching tasks and novel analysis methods using signal detection theory.** A range of reconstruction filters with different behaviours were analysed for a task requiring accurate spatial understanding of a complex scene. In the first experiment detailed in Chapter 5, using simplified data, statistically significant results were found on the accuracy rates between filter choices. The results suggested that reconstruction filters can have an effect on the accuracy results of a path-searching task. Further, stereoscopic presentation of the scene improves the accuracy rates when compared to standard 2D displays.

   Following the results from Chapter 5 an extended experiment was performed in Chapter 6 to test a wider range of reconstruction filters, two transfer functions, and two display modes, using volume data more representative of real volume data sets. Analysing the accuracy rates, the results showed that when more realistic data is used the reconstruction filter has minimal significance, despite visual differences between the generated images. Further, the often quoted post-aliasing and smoothing measurements are not related to task performance. This is natural, as the introduction of Perlin noise reduces sharp boundaries between features, thereby minimising the effect of smoothing and aliasing. The level of opacity was found to have a stronger effect on accuracy, with the opaque renderings leading to higher accuracy rates in most cases. A higher task accuracy was also found when using stereoscopic displays, with this being attributed to a higher number of correct rejections.

   Using techniques from signal detection theory an interaction effect was found between the transfer function and the reconstruction filter with participants having a higher sensitivity to the task when the B-spline filter was combined with opaque rendering. Further, participants were found to produce biased responses towards saying that the nodes were connected with the reconstruction filter making a statistical difference. Participants were also found to be more conservative when the task was presented

in 3D.

These results imply that reconstruction filters can have a more subtle effect on tasks than just affecting overall accuracy and that the frequency domain metrics of the filters cannot be guaranteed to correspond to task performance. The results also show that participants have a higher accuracy in 3D despite producing more conservative results.

3. **A subjective assessment of quality measures of reconstruction filters for stereoscopic volume rendering.**
   Having shown that reconstruction filters can have an effect on path-searching tasks an assessment was performed to determine how observers perceive the differences between stereoscopic DVR images generated with different reconstruction filters. Four quality measures were assessed: depth quality, depth layout, sharpness and minimal jaggyness. The results concluded that reconstruction filters can affect the perceived quality of depth but not the depth layout. This suggests that despite depth quality differing, observers were still able to understand the general layout of the scenes.

   The post-aliasing and smoothing metrics for each filter were found to not have a correlation to any of the measured quality scores. In a novel application of the compactness shape measure, it was found to have a significant strong correlation to each of the quality scores. The correlations suggest that the more compact volumes, such as the skull and engine data set, have higher quality measure scores. We also find that B-spline filter, that produces significant smoothing, has the highest compactness value and should be preferred.

8.3 Future Work

This thesis has presented research into aspects of the DVR pipeline with respect to task performance and observer preference. Given the wide range of parameters for DVR and interaction methods, there are a number of avenues open for further research that could expand upon the work presented in this thesis.

1. **Evaluation of Gradient Estimators in DVR**
   In the DVR pipeline, gradients can be computed at sample points. These gradients

can be used for classification when 2D transfer functions are used, giving more control over the output image. They can also be used for shading, where the gradient is used in place of a surface-normal. As with reconstruction filters, there are a variety of functions that can be used to generate a gradient at a sample point, with the ideal function being the *cosc* function, the derivative of the *sinc* function. Due to the cosc function having unbounded support, practical filters must be used, which due to their non-ideal nature will generate artefacts when used for shading.

Future work would evaluate a range of these gradient estimating functions under task conditions for DVR. This would extend the current understanding of gradient filters and their effect in real-world scenarios. Shading is a depth cue and it can aid the HVS in determining between convex and concave shapes. As different gradient functions create different shading patterns, how these affect the understanding of a scene is a research area that as yet remains unexplored.

2. **Study of Interactive Stereoscopic DVR**

   The experiments performed in this thesis have dealt primarily with static scenes that are not interactive. This has been done to limit the number of variables under investigation. DVR software packages can allow participants to rotate the volume so as to view from different angles. It would be prudent to understand how the parameters in the DVR pipeline affect the performance of tasks when the user is allowed to interact with the volume. It may be that path-searching tasks as used in Chapters 5 and 6 are not suitable, due to them being designed with a static scene in mind. Visual search tasks that require participants to search for specific features that may be visually complex or in a crowded environment could be a suitable alternative and have a precedent of being used in DVR (Laha et al. 2012).

3. **Automatic Assessment of Depth Quality for Stereoscopic DVR**

   The research presented has shown that the parametric choices in the DVR pipeline can impact the quality of depth and Chapter 7 has found that this has a high correlation with the shape compactness measure. Further research should extend the shape compactness measure in order to produce a tool to automatically assess stereoscopic DVR images.

4. **Use of Domain Experts for the Experiments**

   Each of the experiments detailed in the thesis have used novices to perform the tasks.

The reasons for using novices include the difficulty in acquiring a significant number of domain experts as well as reducing prior knowledge as a factor. However, in order to increase the relevance of the tasks and results, future work should elicit the assistance of domain experts for the design of the experiments as well as participating in them.

8.4 Conclusion

In this thesis investigations have been performed into stereoscopic DVR with a focus on whether aspects of the DVR pipeline can impact tasks that require depth perception. Evaluations of the volume resolution and reconstruction filters have been performed for determining whether volume resolution can impact fine depth perception, tasks that require an understanding of the spatial layout of the scene as well as the perceived quality of the stereo DVR images. The results suggest that transfer functions that produce opaque renderings should be preferred and that reconstruction filters can have an effect on task results. We find that the practice of using trilinear interpolation seems justified, given the implementation simplicity, but B-spline approximation could be justified if smoothing is tolerable.

The regularly reported smoothing and post-aliasing properties of reconstruction filters were found to have little correlation with task performance or perceived quality. Instead, the results suggest that image based metrics, such as shape compactness, have a higher correlation to the perceived quality. From subjective evaluations the results suggest that depth quality can vary between parameter choices where as the general understanding of the scene remains mostly unaffected.

In reference to the overall hypothesis of the thesis we found that factors within the DVR pipeline can have an effect on tasks. However, the exact nature of the results can be more subtle than crude effects on accuracy alone. Due to the growing maturity of this field we believe that further progress should focus on the subtle results that require more refined analysis and focused experimental studies.

The results and research presented in the thesis contributes to the understanding of stereoscopic DVR as well as the greater field of 3D depth perception.

Appendix A

Acronyms and Glossary

Acronyms

ANOVA Analysis of Variance.

BCC Body-Centred Cubic.

CC Cartesian Cubic.

CT Computed Tomography.

CT Magnetic Resonance Imaging.

DVR Direct Volume Rendering.

DVRI Direct Volume Rendered Image.

FCC Face-Centred Cubic.

FIR Finite Impulse Response.

GPU Graphics Processing Unit.

HMD Head Mounted Display.

HVS Human Visual System.

IIR Infinite Impulse Response.

IOD Interocular Distance.

KDE Kinetic Depth Effect.

MIP Maximum Intensity Projection.

PEST Parameter Estimation by Sequential Testing.

Glossary

Accommodation Focusing of the lens of the eye on the object of fixation.

Aerial Perspective A depth cue arising from the decrease in sharpness caused by the scattering of light from particles in the atmosphere.

Alias Spectra See stop-band.

Aliasing Caused by sampling below twice the highest frequency of the data set. This introduces spurious high frequencies. In 2D images this introduces stair-case artefacts and sharp edges.

Anaglyph Method of producing 3D images where the images for the left and right eye are filtered by using different coloured filters.

Anisotropy Non-spherically symmetric filters cause asymmetric artefacts in regards to smoothing and post-aliasing.

Autostereoscopic Display A 3D display that does not require viewers to wear headwear to perceive the 3D images.

Band-limited A function is band-limited if it contains no higher frequency than f_{max}, in other words the function has bounded support.

Binocular Depth Cue The depth cue arising from stereopsis where two views of the scene are combined into a single 3D view. This creates a sense of depth.

Binocular Display 3D displays that present a left and right image to the viewer.

Binocular Visual Field Created by the overlapping perspective views from each eye.

Blurring Loss of high-frequencies in the image that relate to the removal of fine details.

Body-Centred Cubic A grid format for a volumetric dataset. Each cell in the lattice contains a sample point at each corner and an extra sample point in the centre of the cell.

Cartesian Cubic Common grid format for a volumetric dataset. The lattice is rectilinear with distances d_x, d_y, d_z between each data sample.

Chromostereopsis The different wavelengths of light are focused in different areas within the eye, this can be used as a depth cue.

Classification Application of a transfer function to a volume. See Transfer Function.

Compactness Measure of how compact a shape is. The disc has the highest compactness equal to 1, where as a shape with branching extremities will attain a low value approaching 0.

Convergence The rotation of the eyes inwards towards a point of focus.

Convolution An operation on a function g and k that produces a new function gk.

Crossed Disparity Occurs when an object appears further than the object fixated on. All objects in front of the horopter will have crossed disparity.

Crosstalk An incomplete separation between the left and right images of a stereoscopic image. This causes light from one image to leak into the other.

Depth Aliasing Artefacts in depth caused by the limited number of steps in depth displayable on a 3D display.

Diplopia When the two views from the eyes cannot be fused and so each view is seen as a separate image.

Direct Volume Rendering A volume rendering method whereby the volumetric data is rendered directly to the image plane. Users have control over the opacity and colour of each of the points in the data set via the use of a transfer function. Implementations of this method include raycasting, splatting and texture mapping.

Disparity Horizontal difference between matching features in the left and right pair of a stereoscopic image. Gives rise to the stereoscopy depth cue.

Domain-based Volume Rendering The volumetric data is transformed into another domain other than spatial before an algorithm used to generate an image.

Extinction Coefficient Function of the number of particles per unit area in a volume when dealing with light participation.

Face-Centred Cubic A grid format for a volumetric dataset. Each cell in the lattice contains a sample point at each corner and a sample point in each face of the cell.

Finite Impulse Response A reconstruction filter whose impulse response is bounded.

Frequency Domain Maps a signal to a range of frequencies.

Fusion When the two views from the eyes are fused into a single 3D scene.

Gradient Estimation A gradient vector is calculated to give an estimate of a surface contained in a volumetric data set.

Horopter The horopter contains all points that are perceived as being at the same depth as the object that is being focused on.

Howard-Dolman Stereo-acuity Test Specific example of a stereo-acuity test. Wooden cylinders were used, for each test one cylinder was moved at a distance from the other cylinder. The participant must then determine which cylinder was closest to themselves. The returned value is the smallest depth that the participant can accurately perceive between the two cylinders.

Image Order Volume Rendering A category of volume rendering algorithms where the algorithm iterates over pixels in the image. See Raycasting.

Indirect Volume Rendering A volume rendering method whereby the volumetric data is mapped to a set of iso-surfaces. These surfaces are then rendered via standard polygon based graphics hardware. The Marching Cubes algorithm is a typical implementation of this method.

Infinite Impulse Response A filter whose impulse response is not bounded.

Inter-perspective Aliasing Aliasing resulting from the finite number of views that can be displayed on a stereoscopic display.

Interocular Distance Distance between the centre of each eye.

Interpolation Scheme See Reconstruction Filter.

Intra-perspective Aliasing The limited horizontal resolution of 3D displays means that there is a limit to the number of steps in depth that can be displayed.

Kinetic Depth Effect A depth cue where the rotation of an object allows viewers to perceive the 3D structure of the object.

Latin-Square Design An n by n grid with n unique values with each specific value appearing once in each row and column. Used to determine the ordering of experiments to prevent learning effects.

Likert A numerical scale used to retrieve the subjective agreement to some given statement by a participant.

Linear Perspective Depth cue that arises from lines converging towards a point in the horizon.

Maximum Intensity Projection An order independent algorithm for producing images from volumetric data. For each ray directed towards the volumetric data only the maximum value along the ray is used.

Monocular Depth Cues Depth cues that can be presented in a 2D image and viewable with a single eye.

Motion Parallax A dynamic depth cue that arises when objects are moving. If two objects are travelling at the same speed, object A will appear to move faster if it is closer whilst object B will appear to move slower if it is further away.

Multi-view Display Autostereoscopic displays that present more than one view of the scene.

Non-correspondence Aliasing Aliasing artefacts specific to stereoscopic images. These occurs when the two images making up the stereoscopic image contain different spatial aliasing artefacts. The different artefacts results in the failure of being able to correctly match points between images.

Nyquist Rate If a function is band-limited then the Nyquist rate is twice the highest frequency f_{max}.

Object Order Volume Rendering A category of volume rendering algorithms where the algorithm iterates over voxels in the volumetric dataset. See Splatting.

Object Space Aliasing Primarily caused when a simple binary voxelisation method is used to create a volumetric data set from a polygonal mesh. As all voxels are either 1 or 0 typical aliasing artefacts are introduced.

Occlusion A depth cue that arises when object obscures all or part of another object.

Oculomotor Depth Cues Depth cues that occur due to the feedback from the muscles in the eyes that used when accommodating or converging on an object.

Order Independent Volume Rendering A category of volume rendering algorithms that do not attempt to calculate the volume rendering integral. Order independent algorithms will produce the same image regardless of the order of projection. See Maximum Intensity Projection.

Parameter Estimation by Sequential Testing An adaptive psychometric test that tracks the number of correct responses from participants and determines the next level the participant should be tested at.

Pass-band The copy of the continuous signal centred at zero.

Path-searching An experiment task where participants must find a path of pre-determined length between two features in a 3D scene.

Perceptual Zooming Occurs when an object that is further away is perceived as being larger than an object of equivalent size that is closer. This happens when perspective projection is not used when creating stereoscopic images.

Perlin Noise An algorithm to produce noise. Used typically in computer graphics to produce pseudo-random visual effects.

Physiological Depth Cues See Oculomotor Depth Cues.

Pictorial Depth Cues See Monocular Depth Cues.

Post-aliasing Aliasing caused during the reconstruction stage of the DVR pipeline. Occurs when a reconstruction filter has non-ideal stop-band behaviour.

Post-classified Transfer Function When using a post-classified transfer function the values in the volume are interpolated and then the colours and opacities in the transfer function are applied.

Pre-classified Transfer Function When using a pre-classified transfer function the colours and opacities are applied to the volume data and then the values are interpolated.

Primary Spectra See pass-band.

Raycasting A DVR algorithm used to generate images from volumetric data. For each pixel in the image a ray is cast towards the volumetric data. Along the ray samples are taken and assigned to colours and opacities.

Reconstruction Filter A function used in DVR algorithms to interpolate between existing data points in a volumtric data set.

Regularity The regularity of a continuous function is the number of times that the function can be differentiated.

Relative Size A depth cue arising from the known relative sizes of two or more objects.

Retinal Disparity Lateral separation between matching points projected onto the retina.

Ringing Oscillations in the output signal near sharp transitions of values in the input signal.

Sample Frequency Ripple Oscillations in the reconstructed volume at the sample frequency.

Sample Lattice Structure defining how the data samples in a volumetric dataset are arranged. Structured layouts include Cartesian Cubic, Body-Centered Cubic and Face-Centered Cubic. Unstructured layouts may have an arbitrary or irregular format.

Sampling Process of taking values from the volumetric data set.

Shading Variation in radiance caused by surface orientation or specular differences.

Shadow Occurs when an object occludes the path to a light source.

Shear Warp A DVR algorithm used to generate images from volumetric data. The volume is treated as a set of slices parallel to the base plane. Rays are then traversed through the slices.

Signal Detection Theory Method of explaining experiments where some visual or auditory signal must be distinguished from a background signal.

Smoothing The removal of high frequencies and rapid variations in the volumetric data during the reconstruction stage of the DVR pipeline. Occurs when a reconstruction filter has non-ideal pass-band behaviour.

Spatial Aliasing Aliasing artefacts specific to 2D images. Occurs when the resolution of the image being displayed is not suitable. Results in jagged edges, discontinuous objects and possible removal of small features.

Splatting A DVR algorithm used to generate images from volumetric data. Each voxel within the volume is projected onto the image plane after being assigned a colour and opacity.

Stereo-acuity Smallest angular difference in depth between points that can be perceived by a viewer.

Stereoblindness Occurs when a person is not able to use the stereopsis depth cue.

Stereoscopic Display A 3D display that requires the viewer to wear some form of passive or active headwear to perceive the 3D image.

Stereoscopic Image An image pair containing a left image and a right image. When viewed correctly they merge to form a 3D image that creates the illusion of depth.

Stereoscopic Resolution The number of steps in depth displayable on a particular 3D display with a given resolution.

Stop-band The alias spectra of the replicated continuous signal not centred at zero.

Texture Gradient A texture or pattern on surface grows closer as the surface recedes into the distance this can be used as a depth cue.

Transfer Function Used to assign colours and opacities to the scalar values in the volumetric data set. Implemented as 1D or 2D lookup table.

Uncrossed Disparity Occurs when an object appears closer than the object fixated on. All object behind the horopter will have uncrossed disparity.

Volume See Volumetric data.

Volume Rendering Method of generating a 2D image from a volumetric data set.

Volume Rendering Integral Equation governing how a ray cast into a volume accumulates colours and opacity.

Volumetric Dataset A set of sample points on a typically 3D grid lattice. Each sample point stores a scalar or vector value that represents some measurement or computed value.

Volumetric Display A display that emits voxels from a true x, y, z position in space.

Voxel 3D analogue to a pixel, typically a single scalar value in a volumetric data set.

Voxelisation A method to convert a standard polygon mesh into a volumetric data set.

Zero-Disparity Plane Plane on a 3D display where objects are displayed with zero-disparity and therefore have the same depth as the screen plane.

Appendix B

Stereoscopic Volume Renderer

This appendix contains an overview of the volume rendering software that was developed for use in the thesis. Each of the experiments detailed in Chapters 4, 5, 6 and 7 used this software to produce the images for the trials.

The software itself was developed by the author of this thesis without the use of standard graphics libraries. This was done both as a learning exercise, to fully understand the volume rendering pipeline and to provide a high-level of flexibility in regards to rendering methods and algorithm parameters.

B.1 Overview

The software produced implements the ray casting method of Levoy (1988) in order to produce images. It is a non-interactive software-based implementation of the algorithm, written primarily in C++ with JavaScript Object Notation (JSON) being used to define the settings files.

By taking advantage of inheritance and dependency injection, the software is straightforward to modify so that extra reconstruction filters, cameras, volume lattices and volume rendering methods can be implemented without requiring substantial changes to the software.

A separate tool, written in Python, is used to define and create the transfer functions that are then passed to the volume renderer as a command-line option.

B.2 Design

The class diagrams and details of parts of the system are given in this section.

In Figure B.1 the class layout for the Camera package used in the volume renderer is given. The camera used is passed as an option in the settings file and is instantiated by the Camera Factory class. The Stereo Camera is a special type of Camera, containing in itself pointers to two other cameras, one for each eye. The Anaglyph and Cross Eye camera differ to the Stereo Camera only in how they output the final image. All options regarding view direction, orientation and image resolution are given in the settings files.

The design of the Interpolation package is given in Figure B.2. Each Interpolator takes a Vector representing the sample position and a Matrix3D that contains the volume data and the double value returned represents the interpolated value. The *preprocess* method is used for interpolation methods that require a preprocessing of the volume data such as the Interpolating B-spline and the Pre-multiplied Trilinear filters.

The design of the Lattice package used to store and interrogate the volume data is given in Figure B.3. Note that for the experiments detailed in the thesis only the CC Lattice has been used, however the BCC and FCC lattices have also been fully implemented. All volumes are loaded as 16-bit unsigned integers, these are then converted to double precision floating point values in the range [0, 1]. The constructor of the Lattice class takes an Interpolator object. When the *intersect* method is called on the Lattice, the position and volume data is passed to the Interpolator to generate an interpolated value between data points in the volume data.

The Point Modifier package displayed in Figure B.4 can be used to alter the colour and opacity of a sample point after it has been classified using the transfer function. For the experiments in the thesis this has been primarily used to introduce Perlin noise into the image, however a Shading Modifier class has also been implemented to add local illumination. Due to the use of the Decorator design pattern the point modifier classes can be chained to allow for more than one modification. For example, Perlin noise can be introduced and then shading can be applied afterwards.

B.3 Ray casting

The software implements Levoy's ray casting method of volume rendering. For each pixel in the image a ray, originating at the camera, is directed into the volume lattice. The

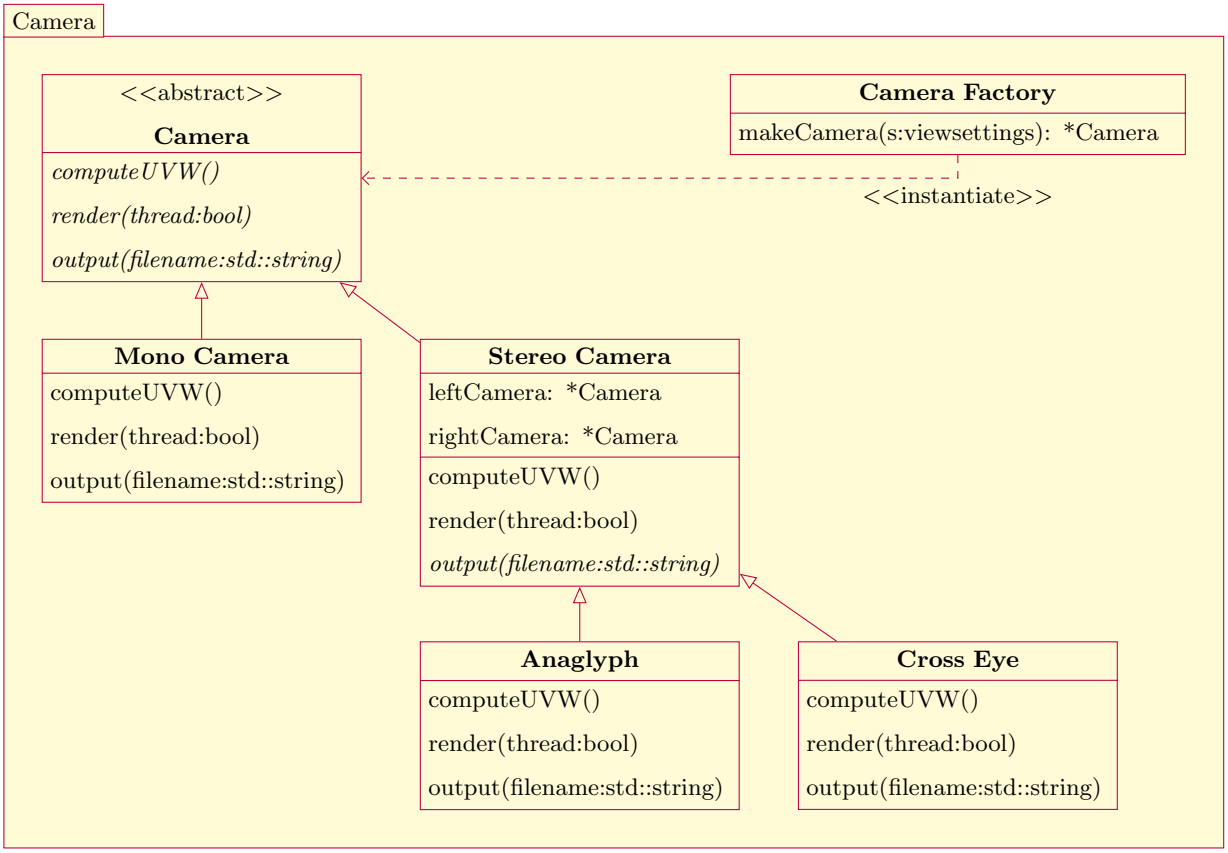


Figure B.1: Class layout of the camera implementation used in the volume renderer.

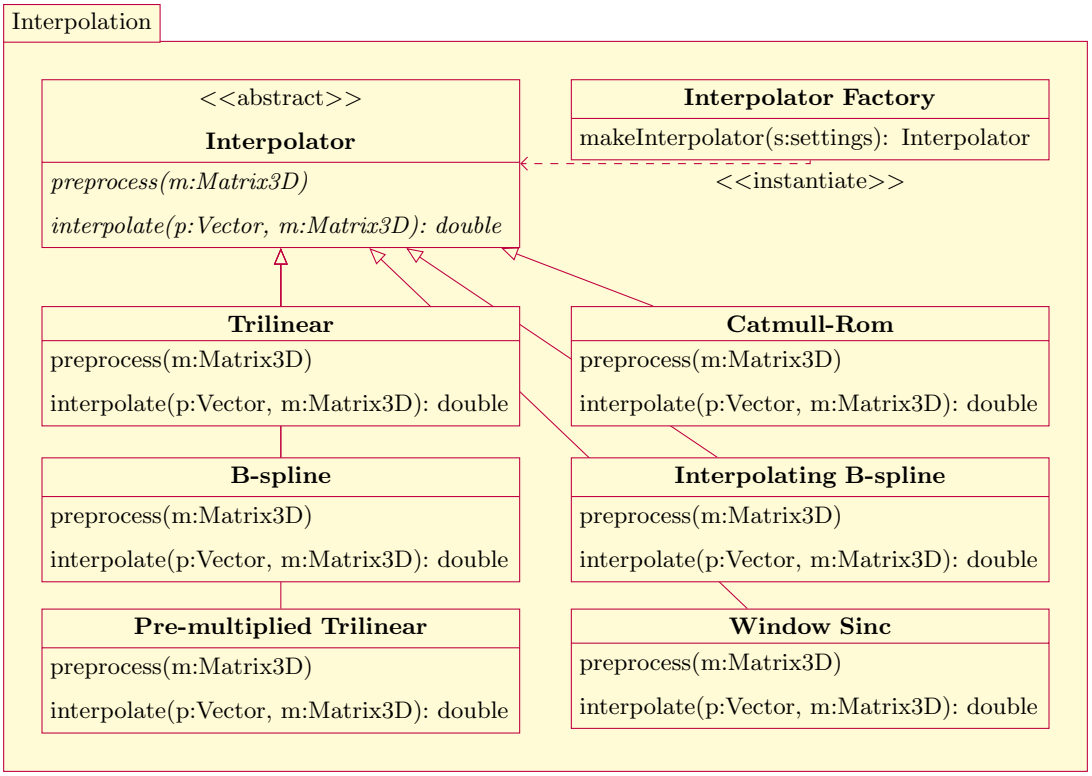


Figure B.2: Class layout of the interpolation package with each of the reconstruction filters implemented.

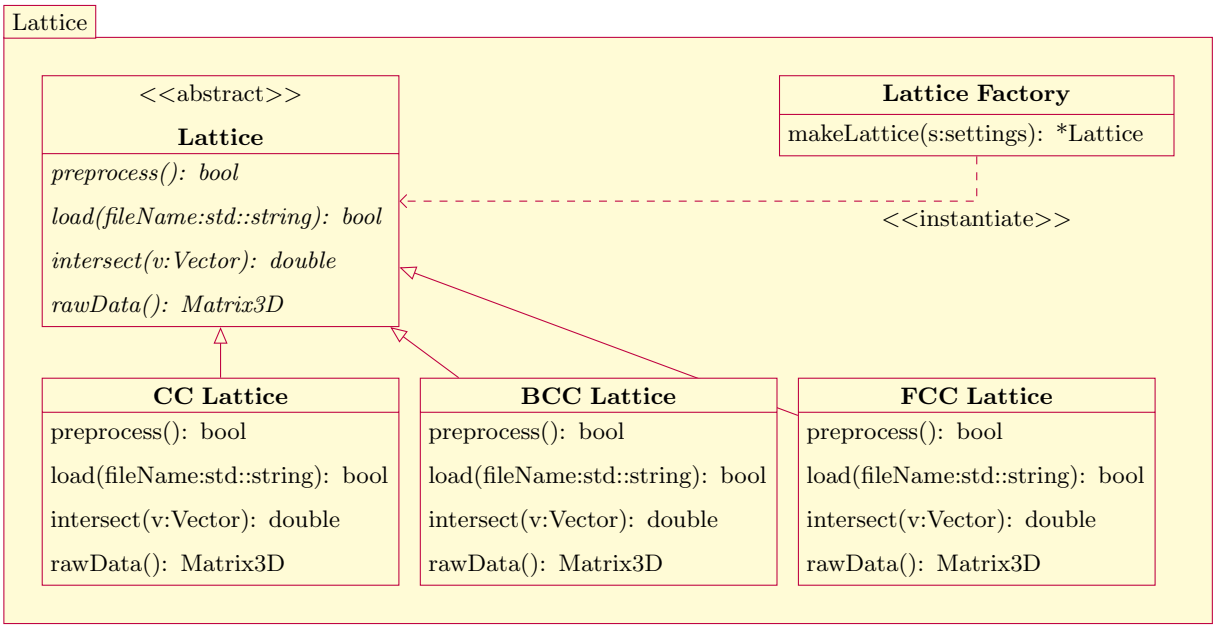


Figure B.3: Class layout of the lattice package used for storing and assessing the volumetric data sets.

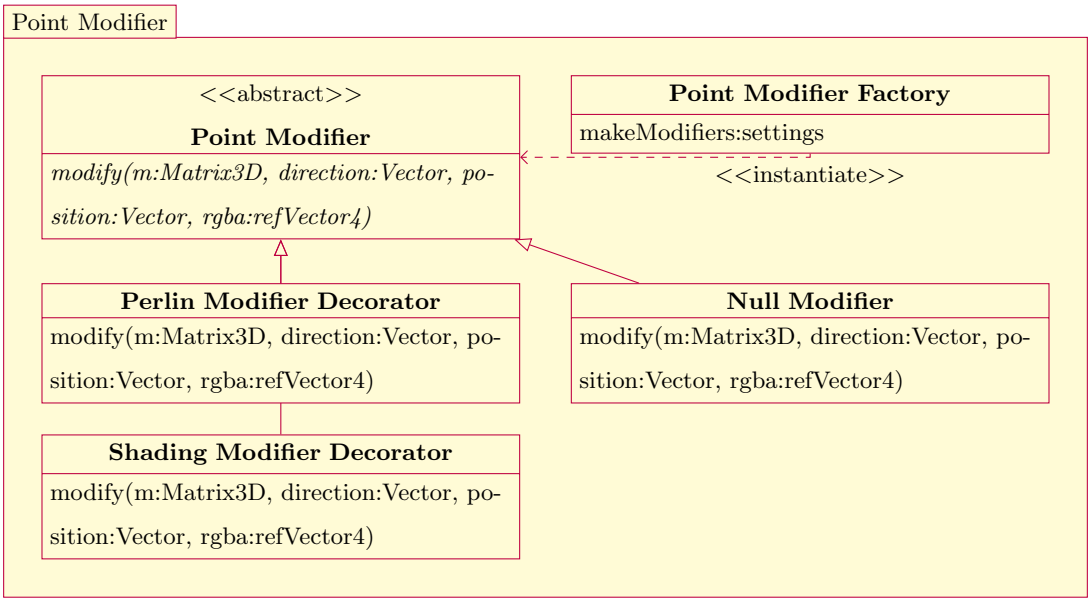


Figure B.4: Class layout of the point modifier package. These are used to modify a sample point after reconstruction to introduce noise or shading.

discrete volume rendering integral is evaluated and the volume data is re-sampled at points along the ray. The distance between samples is fully controllable and is given as an option in the settings file.

The software also supports sub-sampling by firing more than one ray per pixel into the volume. This can be used to reduce the presence of aliasing artefacts at the cost of an increase in computation time. The number of rays to send out per pixel is defined in the settings file.

Although the software is non-interactive, the code does use multi-threading to reduce the processing time required to produce the images. The image is divided into a set of rows, with the exact number being dependent on the number of threads supported by the host hardware. Each thread is given a number of rows to process which are computed using the described ray casting method.

B.4 Transfer Function

The transfer function is loaded as four files, one for each of the RGBa properties. Each file contains a list of 65536 values with each value being a single-precision floating point which are then read into an array of Vectors.

B.5 Settings File

The software is designed to be fully configurable via the JSON settings file. All options, including volume settings, camera options, sampling rate, reconstruction filter and gradient filter are defined in the settings files. Listing B.1 shows an example of a settings file used for the software.

Listing B.1: JSON Settings file

```
{
        "camera": "cross_eye",
        "width": 2400,
        "height": 2400,
        "display_width_px": 1920,
        "display_width_mm": 518.4,
        "distance_to_display_mm": 700,
        "IOD": 65,
        "camera_axis": {"x":0.0, "y":1.0, "z":1.0},
        "camera_angle": 180,
        "sub_sample": false,
        "test_ray": false,
        "sample_step": 0.000488281,
        "subsample_rate": 4,
        "thread": true,
        "background_colour": {"r": 255.0, "g": 255.0, "b":255
           .0, "a":255.0},
        "volume":
        {
                "resolution": {"x":512, "y":512, "z":512},
                "ratio": {"x":1, "y":1, "z":1},
                "file": "../volumes/other/angiography/
                   reference.cc_raw",
                "interpolator": "CC_windowsinc",
                "window_radius": 4,
                "window":"welch",
```

```
25                  "lattice_type": "CC"
26          }
27  }
```

Bibliography

Abhari, K., Baxter, J. S. H., Khan, A. L. I. R. & Peters, T. M. (2015), 'Visual enhancement of MR angiography images to facilitate planning of arteriovenous malformation Interventions', *ACM Transactions on Applied Perception* **12**(1).

Agus, M., Bettio, F., Giachetti, A., Gobbetti, E., Iglesias Guitián, J. A., Marton, F., Nilsson, J. & Pintore, G. (2009), 'An interactive 3D medical visualization system based on a light field display', *The Visual Computer* **25**(9), pp. 883–893.

Albani, L., Marchessoux, C. & Kimpe, T. (2011), 'Character and graphic OLED displays', *Information Display* **27**(10), pp. 24–28.

Anderson, B. L. & Nakayama, K. (1994), 'Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion.', *Psychological review* **101**(3), pp. 414–45.

Andrews, H. (1978), 'Cubic splines for image interpolation and digital filtering', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**(6), pp. 508–517.

Artner, M., Möller, T., Viola, I. & Gröller, M. E. (2005), High-quality volume rendering with resampling in the frequency domain, *in* 'Proceedings of Eurographics / IEEE VGTC Symposium on Visualization', pp. 85–92.

Bentum, M. (1996), Improving image quality of volume-rendered three-dimensional medical data, *in* 'Medical Imaging', Vol. 2707, pp. 32–43.

Beurden, M. H. P. H., IJsselsteijn, W. a. & Juola, J. F. (2012), 'Effectiveness of stereoscopic displays in medicine: A review', *3D Research* **3**(1).

Blake, R. & Sekuler, R. (2006), *Perception*, fifth edn, McGraw-Hill Higher Education, New York.

Blu, T., Thevenaz, P. & Unser, M. (1999), Generalized interpolation: Higher quality at no additional cost, *in* 'Proceedings of IEEE International Conference on Image Processing', number 1, pp. 667–671.

Blu, T., Thévenaz, P. & Unser, M. (2004), 'Linear interpolation revitalized', *IEEE Transactions on Image Processing* **13**(5), pp. 710–719.

Boas, F. E. & Fleischmann, D. (2012), 'CT artifacts : Causes and reduction techniques', *Imaging* **4**, pp. 229–240.

Booth, K. S., Bryden, M. P., Cowan, W. B., Morgan, M. F. & Plante, B. L. (1987), 'On the parameters of human visual performance: an investigation of the benefits of antialiasing', *IEEE Computer Graphics and Applications* **7**, pp. 34–41.

Boucheny, C., Bonneau, G.-P., Droulez, J., Thibault, G. & Ploix, S. (2009), 'A perceptive evaluation of volume rendering techniques', *ACM Transactions on Applied Perception* **5**(4), pp. 1–24.

Breen, D., Mauch, S. & Whitaker, R. (1998), 3D scan conversion of CSG models into distance volumes, *in* 'Proceedings of IEEE Symposium on Volume Visualization', pp. 7–14.

Brenner, D. J. & Hall, E. J. (2007), 'Computed tomography An increasing source of radiation exposure', *New England Journal of Medicine* **357**, pp. 2277–2284.

Carlbom, I. (1993), Optimal filter design for volume reconstruction and visualization, *in* 'Proceedings of IEEE Conference on Visualization', pp. pp. 54–61.

Chelbergy, D., Hsu, J., Babbsz, C., Pizloyy, Z. & Delpy, E. (1994), 'Digital Stereomammography'.

Cho, I., Charlotte, U. N. C., Wartell, Z. & Ribarsky, W. (2012), Evaluating depth perception of volumetric data in semi-immersive VR, *in* 'Proceedings of the International Working Conference on Advanced Visual Interfaces', ACM, pp. 266–269.

Cho, I., Wartell, Z., Dou, W., Wang, X. & Ribarsky, W. (2014), Stereo and motion cueseffect on depth perception of volumetric data, *in* 'Proceedings of SPIE Stereoscopic Displays and Applications'.

Clark, J. J. & Yuille, A. L. (1990), *Data Fusion for Sensory Information Systems*, Kluwer Academic Publishers, Norwell, MA.

Costa, M. F., Moreira, S. M. C. F., Hamer, R. D. & Ventura, D. F. (2010), 'Effects of age and optical blur on real depth stereoacuity.', *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)* **30**(5), pp. 660–6.

Coutant, B. E. & Westheimer, G. (1993), 'Population distribution of stereoscopic ability.', *Ophthalmic Physiological Optics the Journal of the British College of Ophthalmic Opticians Optometrists* **13**(1), pp. 3–7.

Crow, F. C. (1977), 'The aliasing problem in computer-generated shaded images', *Communications of the ACM* **20**(11), pp. 799–805.

Csébfalvi, B. (2008), 'An evaluation of prefiltered reconstruction schemes for volume rendering.', *IEEE Transactions on Visualization and Computer Graphics* **14**(2), pp. 289–301.

Csébfalvi, B. & Domonkos, B. (2009), Interactively controlling the smoothing and postaliasing effects in volume visualization, *in* 'Proceedings of the 25th Spring Conference on Computer Graphics', pp. 129–136.

Cullip, T. & Neumann, U. (1993), Accelerating Volume Reconstruction with 3D Texture Hardware, Technical report, University of North Carolina at Chapel Hill.

Cutting, J. E. & Vishton, P. M. (1995), 'Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth', *Perception* **5**(3), pp. 1–37.

Docherty, A. & Koch, R. (1993), Image distortion in stereoscopic video systems, *in* 'Proceedings of SPIE Symposium on Electronic Imaging: Science and Technology', number 1993, pp. 36–48.

Dodgson, N. (1997), 'Autostereo displays: 3D without glasses', *EID: Electronic Information Displays* **97**(c), pp. 18–20.

Drascic, D. & Milgram, P. (1996), Perceptual issues in augmented reality, *in* 'Proceedings of SPIE Symposium on Electronic Imaging: Science and Technology', pp. 123–134.

Dunne, S., Napel, S. & Rutt, B. (1990), Fast reprojection of volume data, *in* 'Proceedings of the First Conference on Visualization in Biomedical Computing', pp. 11–18.

Elvins, T. (1992), 'A survey of algorithms for volume visualization', *ACM Siggraph Computer Graphics* **Vi**(3).

Engel, K. & Ertl, T. (2002), Interactive high-quality volume rendering with flexible consumer graphics hardware, *in* 'Proceedings of Eurographics'.

Engel, K., Hadwiger, M., Kniss, J., Rezk-Salama, C. & Weiskopf, D. (2006), *Real-Time Volume Graphics*, A K Peters/CRC Press.

Favalora, G. (2005), 'Volumetric 3D displays and application infrastructure', *Computer* pp. 37–44.

Foley, J. D., Van Dam, A., Feiner, S. K. & Hughes, J. F. (1990), *Computer Graphics: Principles and Practice*, Vol. 2nd of *The Systems Programming Series*, Addison-Wesley.

Gibson, E. J., Gibson, J. J., Smith, O. W. & Flock, H. (1959), 'Motion parallax as a determinant of perceived depth.', *Journal of Experimental Psychology* **58**(1), pp. 40–51.

Gibson, J. (1950), *The perception of the visual world.*, Houghton Mifflin.

Glassner, A. (1995), *Principles of Digital Image Synthesis*, Morgan Kaufmann.

Goldstein, E. (2002), *Sensation and Perception*, sixth edn, Wadsworth.

Grosset, P. & Schott, M. (2013), 'Evaluation of Depth of Field for Depth Perception in DVR', *IEEE Pacific Visualization Symposium* **2013**, pp. 1–8.

Haddad, R. A. & Parsons, T. W. (1991), *Digital signal processing : Theory, applications, and hardware*, Electrical engineering communications and signal processing series, Computer Science Press.

Halpern, D. L. & Blake, R. R. (1988), 'How contrast affects stereoacuity.', *Perception* **17**(4), pp. 483–495.

Hancock, D. (2001), Distributed volume rendering and stereoscopic display for radiotherapy treatment planning, Phd, University of Manchester.

Hassaine, D., Holliman, N. S. & Liversedge, S. P. (2010), 'Investigating the performance of path-searching tasks in depth on multiview displays', *ACM Transactions on Applied Perception* **8**(1), pp. 1–18.

He, T., Hong, L., Kaufman, A. & Pfister, H. (1996), Generation of transfer functions with stochastic search techniques, *in* 'Visualization '96. Proceedings.', pp. 227–234.

Hecht, S. & Smith, E. L. (1936), 'Intermittent stimulation by light : Vi. Area and the relation between critical frequency and intensity.', *The Journal of General Physiology* **19**(6), 979–89.

Hodges, L. & Davis, E. (1993), 'Geometric considerations for stereoscopic virtual environments', *Presence* pp. 34–43.

Holliman, N. (2005), 3D Display Systems, *in* J. Dakin & R. Brown, eds, 'Handbook of Optoelectronics', Taylor & Francis, New York and London, pp. pp. 1067–1100.

Holliman, N. (2011), 'Three-dimensional displays: a review and applications analysis', *IEEE Transactions on Broadcasting* **57**(2), 362–371.

Howard, H. (1919), 'A test for the judgement of distance', *American Journal of Opthalmology* pp. 656–675.

Hsu, J. & Babbs, C. F. (1993), 'Pre-clinical ROC studies of digital stereomammography', *IEEE Transactions on Medical Imaging* **14**(2), pp. 318–327.

Hsu, J., Babbs, C. F., Chelberg, D. M., Pizlo, Z. & Delp, E. J. (1993), A study of the effectiveness of stereo imaging with applications in mammography, *in* 'Proceedings of SPIE Symposium on Electronic Imaging: Science and Technology', pp. 154–165.

Hsu, J. & Chelberg, D. (1993), 'Visible light and X-ray ray tracing of generalized cylinders', *ECE Technical Reports* .

Hubbold, R., Hancock, D. & Moore, C. (1997), Autostereoscopic display for radiotherapy planning, *in* 'Proceedings of Electronic Imaging', pp. 16–27.

Hubbold, R., Hancock, D. & Moore, C. (1998), Stereoscopic volume rendering, *in* 'Proceedings of Visualisation in Scientific Computing', pp. 105–115.

Ichihara, S., Kitagawa, N. & Akutsu, H. (2007), 'Contrast and depth perception: Effects of texture contrast and area contrast', *Perception* **36**(5), pp. 686–695.

ITU-R (2002), 'Methodology for the subjective assessment of the quality of television pictures', *Technical Report. Recommendation ITU-R BT.500-11* **13**, pp. 1–48.

Jää-Aro, K. & Kjelldahl, L. (1997), Effects of image resolution on depth perception in stereo and nonstereo images, *in* 'Proceedings of Electronic Imaging', pp. 319–326.

John J. Wild, J. M. R. (1952), 'Application of echo-ranging techniques to the determination of structure of biological tissues', *Science* **115**(2983), 226–230.

Johnson, L. G., Edwards, P. & Hawkes, D. (2003), 'Surface transparency makes stereo overlays unpredictable: the implications for augmented reality', *Studies in Health Technology and Informatics* **94**, pp. 131–136.

Johnston, E. B., Cumming, B. G. & Landy, M. S. (1994), 'Integration of stereopsis and motion shape cues.', *Vision research* **34**(17), pp. 2259–2275.

Jones, G., Lee, D., Holliman, N. & Ezra, D. (2001), Controlling perceived depth in stereoscopic images., *in* 'SPIE Stereoscopic Displays and Virtual Reality Systems VIII', pp. 42–53.

Jones, M. W., Baerentzen, J. A. & Sramek, M. (2006), '3d distance fields: a survey of techniques and applications', *IEEE Transactions on Visualization and Computer Graphics* **12**(4), 581–599.

Jones, M. W. & Satherley, R. (2000), Voxelisation: Modelling for volume graphics, *in* 'Vision, Modeling, and Visualisation 2000', Press, pp. 319–326.

Julesz, B., Papathomas, T. V. & Phillips, F. (2006), *Foundations of Cyclopean Perception*, MIT Press.

Jung, M., Park, H. & Paik, D. (1998), An analytical ray casting of volume data, *in* 'Proceedings of Sixth Pacific Conference on Computer Graphics and Applications', pp. 76–86.

Juweid, M. E. & Cheson, B. D. (2006), 'Positron-emission tomography and assessment of cancer therapy', *New England Journal of Medicine* **354**(5), 496–507.

Kaufman, A. E. (1987), Algorithm for 3D scan-conversion of polygons, *in* 'Proceedings of Eurographics', pp. 197–208.

Kaufman, A. & Mueller, K. (2005), 'Overview of volume rendering', *The Visualization Handbook* **7**, pp. 127–174.

Kersten, M. (2006), Stereoscopic volume rendering of medical images, Masters, Queen's University.

Kersten, M. a., Stewart, a. J., Troje, N. & Ellis, R. (2006), 'Enhancing depth perception in translucent volumes.', *IEEE Transactions on Visualization and Computer Graphics* **12**(5), pp. 1117–1124.

Kersten-Oertel, M., Chen, S. J.-S. & Collins, D. L. (2014), 'An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery.', *IEEE Transactions on Visualization and Computer Graphics* **20**(3), pp. 391–403.

Keys, R. (1981), 'Cubic convolution interpolation for digital image processing', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **29**, pp. 1153–1160.

Kindlmann, G. & Durkin, J. (1998), 'Semi-automatic generation of transfer functions for direct volume rendering', *Proceedings of the 1998 IEEE Symposium on Volume Visualization* pp. pp. 79–86.

Kruger, J. & Westermann, R. (2003), Acceleration techniques for GPU-based volume rendering, *in* 'Proceedings of the 14th IEEE Visualization', p. pp. 38.

Künsch, H. R., Agrell, E. & Hamprecht, F. A. (2005), 'Optimal lattices for sampling', *IEEE Transactions on Information Theory* **51**(2), pp. 634–647.

Lacroute, P. & Levoy, M. (1994), Fast volume rendering using a shear-warp factorization of the viewing transformation, *in* 'Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques', pp. 451–458.

Laha, B., Bowman, D. a. & Socha, J. J. (2014), 'Effects of VR system fidelity on analyzing isosurface visualization of volume datasets', *IEEE Transactions on Visualization and Computer Graphics* **20**(4), pp. 513–522.

Laha, B., Sensharma, K., Schiffbauer, J. D. & Bowman, D. a. (2012), 'Effects of immersion on visual analysis of volume data.', *IEEE Transactions on Visualization and Computer Graphics* **18**(4), pp. 597–606.

LaMar, E., Hamann, B. & Joy, K. (1999), Multiresolution techniques for interactive texture-based volume visualisation, *in* 'Proceedings of IEEE Visualization', pp. 365–374.

Lambooij, M. (2007), Visual discomfort in stereoscopic displays: a review, *in* 'Electronic Imaging', pp. 1–13.

Lambooij, M. (2011), 'Evaluation of stereoscopic images: beyond 2D quality', *IEEE Transactions on Broadcasting* **57**(2), 432–444.

Landy, M. & Johnston, E. (1995), 'Measurement and modeling of depth cue combination : In defense of weak fusion', *Vision Research* **35**(3), pp. 389–412.

Larson, W. & Bolduc, M. (1991), 'Effect of induced blur on visual acuity and stereoacuity', *Optometry & Vision Science* .

Lebreton, P., Raake, A., Barkowsky, M. & Le Callet, P. (2012), 'Evaluating depth perception of 3D stereoscopic videos', *IEEE Journal of Selected Topics in Signal Processing* **6**(6), pp. 710–720.

Lee, C.-h., Koo, Y.-m. & Shin, Y. G. (1997), 'Template-based rendering of run-length encoded volumes', *Proceedings The Fifth Pacific Conference on Computer Graphics and Applications* **i**, 138–147.

Legge, G. & Yuanchao, G. (1989), 'Stereopsis and Contrast', *Vision Research* **29**(8), pp. 989–1004.

Levoy, M. (1987), 'Rendering of surfaces from volumetric data', *Report TR87-016, Chapel Hill, University of North Carolina* **1987**.

Levoy, M. (1988), 'Display of surfaces from volume data', *IEEE Computer Graphics and Applications* **8**(3), pp. 29–37.

Levoy, M. (1990), 'Efficient ray tracing of volume data', *ACM Transactions on Graphics* **9**(3), 245–261.

Levoy, M. (1992), Volume rendering using the fourier projection-slice theorem, *in* 'Proceedings of the Conference on Graphics Interface', Computer Systems Laboratory, Stanford University, pp. 61–69.

Li, W., Goodchild, M. F. & Church, R. (2013), 'An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems', *International Journal of Geographical Information Science* **27**(6), pp. 1227–1250.

Lipscomb, J. (1989), Experience with stereoscopic display devices and output algorithms, *in* 'Proceedings of SPIE Three-Dimensional Visualization and Display Technologies', pp. 28–34.

Lipton, L. (1997), 'Stereographics Developers' Handbook', *StereoGraphics Corporation* .

Lorensen, W. & Cline, H. (1987), 'Marching cubes: A high resolution 3D surface construction algorithm', *ACM Siggraph Computer Graphics* **21**(4), pp. 163–169.

Machiraju, R., Swan, E. & Yagel, R. (1995), 'Spatial domain characterization and control of reconstruction errors', *Rendering Techniques 1995* pp. pp. 64–73.

Macmillan, N. A. & Douglas Creelman, C. (2004), *Detection Theory: A User's Guide*, 2, revised edn, Taylor & Francis.

Malzbender, T. (1993), 'Fourier volume rendering', *ACM Transactions on Graphics* **12**(3), pp. 233–250.

Marks, J., Andalman, B., Beardsley, P. A., Freeman, W., Gibson, S., Hodgins, J., Kang, T., Mirtich, B., Pfister, H., Ruml, W., Ryall, K., Seims, J. & Shieber, S. (1997), Design galleries: A general approach to setting parameters for computer graphics and animation, *in* 'Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques', SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp. 389–400.

Marschner, S. & Lobb, R. (1994), An evaluation of reconstruction filters for volume rendering, *in* 'Proceedings of the Conference on Visualization', IEEE Computer Society Press, pp. 100–107.

Martinez Escobar, M., Juhnke, B., Hisley, K., Eliot, D. & Winer, E. (2013), Assessment of visual-spatial skills in medical context tasks when using monoscopic and stereoscopic visualization, *in* C. R. Mello-Thoms, ed., 'Proceedings of SPIE Medical Imaging', pp. 86730N—-86730N.

Mather, G. & Smith, D. R. R. (2002), 'Blur discrimination and its relation to blur-mediated depth perception', *Perception* **31**(10), pp. 1211–1219.

Max, N. (1995), 'Optical models for direct volume rendering', *IEEE Transactions on Visualization and Computer Graphics* **1**(2), pp. 99–108.

Mcallister, D. F., ed. (1993), *Stereo computer graphics and other true 3D technologies*, Princeton University Press.

Meißner, M., Huang, J., Bartz, D., Mueller, K. & Crawfis, R. (2000), A practical evaluation of popular volume rendering algorithms, *in* 'Proceedings of the 2000 IEEE Symposium on Volume Visualization', ACM Press, New York, New York, USA, pp. 81–90.

Michelson, A. A. (1962), *Studies in Optics*, Dover Publications.

Mitchell, D. P. & Netravali, A. N. (1988), 'Reconstruction filters in computer-graphics', *ACM SIGGRAPH Computer Graphics* **22**(4), pp. 221–228.

Moller, C. N. & Travis, A. R. L. (2005), 'Correcting interperspective aliasing in autostereoscopic displays.', *IEEE Transactions on Visualization and Computer Graphics* **11**(2), pp. 228–236.

Moller, T. & Machiraju, R. (1997), A comparison of normal estimation schemes, *in* 'Proceedings Visualization', pp. 19–26.

Möller, T., Machiraju, R., Mueller, K. & Yagel, R. (1996), Classification and local error estimation of interpolation and derivative filters for volume rendering, *in* 'Proceedings of the 1996 Symposium on Volume Visualization', pp. 71–78.

Montgomery, D. C. (2009), *Design and Analysis of Experiments*, seventh edn, John Wiley & Sons.

Mora, B. & Evert, D. (2004), 'Instant volumetric understanding with order independent volume rendering', *Computer Graphics Forum* **23**(3).

Moreland, K. (2004), Fast high accuracy volume rendering, Phd, University of New Mexico.

Mueller, K., Moller, T. & Crawlis, R. (1999), 'Splatting without the blur', *Proceedings Visualization '99* pp. 363–544.

Mulder, J. D. & van Liere, R. (2000), 'Fast perception-based depth of field rendering', *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* pp. 129–133.

Nagata, S. (1984), How to reinforce perception of depth in single two-dimensional pictures, *in* 'Proceedings of the Society for Information Display', Vol. 25, pp. 239–246.

Nelson, T., Ji, E. & Lee, J. (2008), 'Stereoscopic evaluation of fetal bony structures', *American Institute of Ultrasound in Medicine* **d**, pp. 15–24.

Nielsen, K. & Poggio, T. (1984), 'Vertical image registration in stereopsis', *Vision Research* **24**(10), pp. 1133–1140.

Novins, K. (1993), Towards accurate and efficient volume rendering, Technical report, Cornell University.

Novotny, P., Dimitrov, L. & Sramek, M. (2010), 'Enhanced voxelization and representation of objects with sharp details in truncated distance fields', *IEEE Transactions on Visualization and Computer Graphics* **16**(3), pp. 484–498.

Odell, N. V., Hatt, S. R., Leske, D. a., Adams, W. E. & Holmes, J. M. (2009), 'The effect of induced monocular blur on measures of stereoacuity.', *Journal of AAPOS : The Official Publication of the American Association for Pediatric Ophthalmology and Strabismus / American Association for Pediatric Ophthalmology and Strabismus* **13**(2), pp. 136–141.

Ollinger, J. M. & Fessler, J. A. (1997), 'Positron-emission tomography', *IEEE Signal Processing Magazine* **14**(1), 43–55.

O'Shea, R. P., Govan, D. G. & Sekuler, R. (1997), 'Blur and contrast as pictorial depth cues.', *Perception* **26**(5), pp. 599–612.

Osserman, R. (1978), 'The isoperimetric inequality', *Bulletin of the American Mathematical Society* **84**(6), pp. 1182–1239.

Otsu, N. (1979), 'A threshold selection method from gray-level histograms', *IEEE Transactions on Systems, Man, and Cybernetics* **9**(1), pp. 62–66.

Palmer, S. E. (1999), *Vision Science: Photons to Phenomenology*, Vol. 4, MIT Press.

Pantaleoni, J. (2011), VoxelPipe : A programmable pipeline for 3D voxelization blending-based rasterization, *in* 'Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics', Vol. 1, pp. 99–106.

Parrott, R. & Stytz, M. (1993), 'Towards statistically optimal interpolation for 3D medical imaging', *IEEE Engineering in Medicine and Biology Magazine* **12**(3), pp. 49–59.

Pastoor, S. (1995), Human factors of 3D imaging: Results of recent research at Heinrich-Hertz-Institut Berlin, *in* 'Proceedings of the International Display Workshop', Vol. 95, Asian Technology Information Program, pp. 66–72.

Pastoor, S. & Wöpking, M. (1997), '3-D displays: A review of current technologies', *Displays* **17**.

Pentland, a. P. (1987), 'A new sense for depth of field.', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**(4), pp. 523–531.

Pfautz, J. D. (2000), Depth Perception in Computer Graphics, PhD thesis.

Pickens, D. (2000), Magnetic Resonance Imaging, *in* 'Handbook of Medical Imaging, volume 1: Physics and Psychophysics', SPIE Press Bellingham, Washington, chapter 6, pp. 373–461.

Porter, T. & Duff, T. (1984), 'Compositing Digital Images', *ACM Siggraph Computer Graphics* **18**(3), pp. 253–259.

Poston, T., Serra, L., Lawton, W. & Chua, B. (1995), Interactive tube finding on a virtual workbench, *in* 'Proceedings of the Second International Symposium on Medical Robotics and Computer Assisted Surgery', pp. 119–123.

Prabhat, Forsberg, A., Katzourin, M., Wharton, K. & Slater, M. (2008), 'A comparative study of desktop , fishtank , and cave systems for the exploration of volume rendered confocal data sets', *IEEE Transactions on Visualization and Computer Graphics* **14**(3), pp. 551–563.

Prauchner, J. L., Freitas, C. M. D. S. & Comba, J. L. D. (2005), Two-level interaction approach for transfer function specification, *in* 'XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05)', pp. 265–272.

Ramachandran, V. (1988), 'Perceiving shape from shading', *Scientific American* **259**(2), pp. 76–83.

Reichelt, S., Häussler, R., Fütterer, G. & Leister, N. (2010), Depth cues in human visual perception and their realization in 3D displays, *in* B. Javidi, J.-Y. Son, J. T. Thomas & D. D. Desjardins, eds, 'Proceedings of SPIE Defense, Security and Sensing', pp. 76900B–76900B–12.

Reinhart, W. (1992), Gray-scale requirements for anti-aliasing of stereoscopic graphic imagery, *in* 'Proceedings of SPIE Symposium on Electronic Imaging: Science and Technology', pp. 90–100.

Rezk-Salama, C. & Engel, K. (2000), Interactive volume rendering on standard PC graphics hardware using multi-textures and multi-stage rasterization, *in* 'Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Workshop on Graphics Hardware', pp. 109–118.

Richards, W. (1970), 'Stereopsis and stereoblindness.', *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale* **10**(4), pp. 380–388.

Richards, W. & Foley, J. M. (1974), 'Effect of luminance and contrast on processing large disparities', *Journal of the Optical Society of America* **64**(12), pp. 1703.

Richards, W. & Miller, J. F. (1969), 'Convergence as a cue to depth', *Perception & Psychophysics* **5**(5), pp. 317–320.

Roberts, D. A. T. (2014), Investigating depth perception with stereoscopic volume rendering, *in* 'Theory and Practice of Computer Graphics', pp. 1–2.

Roberts, D. A. T. & Ivrissimtzis, I. (2016*a*), 'Quality measures of reconstruction filters for stereoscopic volume rendering', *Computational Visual Media* **2**(1), pp. 1–12.

Roberts, D. A. T. & Ivrissimtzis, I. (2016*b*), 'Reevaluating reconstruction filters for path-searching tasks in 3d', *Computer Graphics Forum* . doi:10.1111/cgf.12939.

Roberts, D. A. T., Ivrissimtzis, I. & Holliman, N. S. (2014), An evaluation of reconstruction filters for a path-searching task in 3D, *in* 'Sixth International Workshop on Quality of Multimedia Experience', pp. 165–170.

Rogers, B. & Graham, M. (1979), 'Motion parallax as an independent cue for depth perception', *Perception* .

Rohaly, A. M. & Wilson, H. R. (1999), 'The effects of contrast on perceived depth and depth discrimination.', *Vision Research* **39**(1), pp. 9–18.

Ropinski, T., Doring, C. & Rezk-Salama, C. (2010), 'Interactive volumetric lighting simulating scattering and shadowing', *2010 IEEE Pacific Visualization Symposium (PacificVis)* pp. 169–176.

Ropinski, T., Steinicke, F. & Hinrichs, K. (2006), 'Visually supporting depth perception in angiography imaging', *Smart Graphics* pp. 93–104.

Schlegel, P. & Pajarola, R. (2013), 'Visibility-difference entropy for automatic transfer function generation'.

Schmidt, P. P. (1994), 'Sensitivity of random dot stereoacuity and Snellen acuity to optical blur.', *Optometry and Vision Science Official Publication of the American Academy of Optometry* **71**(7), pp. 466–471.

Schreiber, W. F. & Troxel, D. E. (1985), 'Transformation between continuous and discrete representations of images: A perceptual approach', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-7**(2), pp. 178–186.

Schroeder, W. & Martin, K. (2001), 'The transfer function bake-off', *IEEE Computer Graphics and Applications* **21**(3), pp. 16–22.

Schwarz, M. & Seidel, H.-P. (2010), 'Fast parallel surface and solid voxelization on GPUs', *ACM Transactions on Graphics* **29**(6), pp. 179.

Seuntiens, P., Meesters, L. & Ijsselsteijn, W. (2006), 'Perceived quality of compressed stereoscopic images', *ACM Transactions on Applied Perception* **3**(2), pp. 95–109.

Sexton, I. & Surman, P. (1999), 'Stereoscopic and autosterescopic display systems', *IEEE Signal Processing Magazine* **16**(3), pp. 85–99.

Shannon, C. (1949), 'Communication in the presence of noise', *Proceedings of the IRE* **37**(1), pp. 10–21.

Sigg, C. & Hadwiger, M. (2005), 'Fast third-order texture filtering', *GPU gems* **2**, pp. 313–329.

Soret, M., Bacharach, S. L. & Buvat, I. (2007), 'Partial-volume effect in pet tumor imaging', *Journal of Nuclear Medicine* **48**(6), 932–945.

Souza, A., Udupa, J. K. & Saha, P. K. (2005), 'Volume rendering in the presence of partial volume effects', *IEEE Transactions on Medical Imaging* **24**(2), 223–235.

Spitzer, V. M. & Whitlock, D. G. (1998), 'The visible human dataset: The anatomical platform for human simulation'.

Sramek, M. & Kaufman, A. (1998), 'Object voxelization by filtering', *IEEE Symposium on Volume Visualization* pp. 111–118.

Sramek, M. & Kaufman, A. (1999), 'Alias-free voxelization of geometric objects', *IEEE Transactions on Visualization and Computer Graphics* **5**(3), pp. 251–267.

Stigmar, G. (1971), 'Blurred visual stimuli', *Acta Ophthalmologica* **49**(3), pp. 364–379.

Tam, W. J. (1998), Psychovisual aspects of viewing stereoscopic video sequences, *in* M. T. Bolas, S. S. Fisher & J. O. Merritt, eds, 'Proceedings of Electronic Imaging', Vol. 3295, SPIE, pp. 226–235.

Taylor, M. & Creelman, C. (1967), 'PEST : Efficient estimates on probability functions', *The Journal of the Acoustical Society of America* **41**(4).

Theul, T. (1999*a*), 'On Windowing for Gradient Estimation in Volume Visualization', **Vi**.

Theul, T. (1999*b*), Sampling and reconstruction in volume visualization, Phd, Vienna University of Technology.

Theußl, T., Hauser, H. & Gröller, E. (2000), Mastering windows: Improving reconstruction, *in* 'Proceedings of the 2000 IEEE Symposium on Volume Visualization', pp. 101–108.

Thévenaz, P., Blu, T. & Unser, M. (2000*a*), 'Image interpolation and resampling', *Handbook of Medical Imaging, Processing and Analysis* pp. 393–420.

Thévenaz, P., Blu, T. & Unser, M. (2000*b*), 'Interpolation revisited.', *IEEE Transactions on Medical Imaging* **19**(7), pp. 739–758.

Totsuka, T. & Levoy, M. (1993), 'Frequency domain volume rendering', *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* pp. 271–278.

Tzeng, F. Y., Lum, E. B. & Ma, K. L. (2005), 'An intelligent system approach to higher-dimensional classification of volume data', *IEEE Transactions on Visualization and Computer Graphics* **11**(3), 273–284.

Ukai, K. & Howarth, P. A. (2008), 'Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations', *Displays* **29**(2), 106–116.

van Beurden, M. (2010), 'Performance of a path tracing task using stereoscopic and motion based depth cues', *Quality of Multimedia Experience* pp. 176–181.

van Beurden, M. H., Ijsselsteijn, W. a. & a.W. de Kort, Y. (2011), Evaluating stereoscopic displays: Both efficiency measures and perceived workload sensitive to manipulations in binocular disparity, *in* A. J. Woods, N. S. Holliman & N. A. Dodgson, eds, 'Proceedings of SPIE Electronic Imaging', pp. 786316–786316.

van Dijk, A. M., Martens, J.-B. & Watson, A. B. (1995), Quality asessment of coded images using numerical category scaling, *in* 'Proceedings of SPIE Advanced Image and Video Communications and Storage Technologies', Vol. 2451, pp. 90–101.

Wang, S. & Kaufman, A. (1993), Volume sampled voxelization of geometric primitives, *in* 'Proceedings of IEEE Conference on Visualization', pp. 78–84.

Wang, W. & Kaufman, A. E. (1994), 'Volume-sampled 3D modeling', *IEEE Computer Graphics and Applications* **14**(5), pp. 26–32.

Ware, C. (2004), *Information Visualization: Perception for Design*, Vol. 22 of *The Morgan Kaufmann series in interactive technologies / ed. Stuart Card. - San Francisco : Morgan Kaufmann*, Morgan Kaufmann.

Ware, C. & Franck, G. (1996), 'Evaluating stereo and motion cues for visualizing information nets in three dimensions', *ACM Transactions on Graphics* **15**(2), pp. 121–140.

Ware, C. & Mitchell, P. (2005), 'Reevaluating stereo and motion cues for visualizing graphs in three dimensions', *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* pp. 51–58.

Westheimer, G. & McKee, S. (1980), 'Stereoscopic acuity with defocused and spatially filtered retinal images', *Journal of the Optical Society of America* **70**, pp. 772–778.

Westover, L. (1990), Footprint evaluation for volume rendering, *in* 'Proceedings of the ACM SIGGRAPH Computer Graphics', Vol. 24, pp. 367–376.

Wheatstone, C. (1838), 'Contributions to the physiology of vision.–Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision', *Philosophical Transactions of the Royal Society of London* **128**, pp. 371–394.

Wilcox, L., Elder, J. & Hess, R. (2000), 'The effects of blur and size on monocular and stereoscopic localization', *Vision Research* **40**(26), pp. 3575–3584.

Williams, P. L. & Uselton, S. P. (1999), 'Metrics and generation specifications for comparing volume-rendered images', *The Journal of Visualization and Computer Animation* **10**(3), pp. 159–178.

Yagel, R. & Ciula, K. (1994), High quality template-based volume rendering, Technical report, Ohio State University.

Yagel, R. & Kaufman, A. (1992), 'Template-based volume viewing', *Computer Graphics Forum* **11**(3), pp. 153–167.

Yeh, Y. Y. & Silverstein, L. D. (1990), 'Limits of fusion and depth judgment in stereoscopic color displays.', *Human Factors* **32**(1), pp. 45–60.

Zhang, J.-W. & Sun, J.-Z. (2003), Adaptive transfer function design for volume renering by using a general regression neural network [renering read rendering], *in* 'Machine Learning and Cybernetics, 2003 International Conference on', Vol. 4, pp. 2234–2239 Vol.4.

Zheng, Z., Xu, W. & Mueller, K. (2010), 'VDVR: verifiable visualization of projection-based data.', *IEEE Transactions on Visualization and Computer Graphics* **16**(6), pp. 1515–1524.